Identification of Interacting Partner(s) of SARS-CoV Spike Glycoprotein

CHUCK Chi-pang

B.Sc. (Hons). CUHK

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy
In
Biochemistry

September 2006

I declare that the assignment here submitted is original except for source material explicitly acknowledged. I also acknowledge that I am aware of University policy and regulations on honesty in academic work, and of the disciplinary guidelines and procedures applicable to breaches of such policy and regulations, as contained in the website http://www.cuhk.edu.hk/policy/academichonesty/

[signature]	2006/09/15
Signature	Date
Chuck Chi Pang	04073460
Name	Student ID

BCH 806R	Thesis Research
Course code	Course title

THESIS COMMITTEE

Professor WONG Kam Bo (Chair)

Professor TSUI Kwok Wing, Stephen (Thesis Supervisor)

Professor AU Wing Ngor, Shannon (Committee Member)

Professor ZHU Guang (External Examiner)

ABSTRACT

Severe Acute Respiratory Syndrome (SARS) is a novel disease that induced a global tragedy in 2003. Totally, 916 human deaths were resulted and more than U.S. $30 billions were lost worldwide. The pathogen was discovered and named SARS-associated coronavirus (SARS-CoV). The spike glycoprotein, one of the four known structural proteins, is related to viral entry, host range, tissue tropism and virulence. This glycoprotein has two functional subunits S1 and S2, and they are responsible for receptor-binding and membrane fusion respectively. Nonetheless, the viral entry mechanism is not yet fully understood. Identification of the interacting partners might be helpful to find out the mechanism of the SARS pathogenesis and, in turn, develop an effective neutralizing vaccine against SARS-CoV.

At present, two functional receptors, angiotensin-converting enzyme 2 (ACE2) and CD209L, are identified. Tissue distributions of these receptors are however different from that of SARS-CoV. Some tissues highly expressing ACE2 and CD209L are insusceptible. So, we believe that other unidentified receptor(s) or co-receptor(s) remain to be discovered.

In this study, various DNA fragments encoding truncated spike proteins were cloned into pGEX-6P-1, pET-28a, pRHisMBP and pPICZα-A. The truncated proteins were then expressed in *E. coli* and *P. pastoris* expression systems. After optimizing the conditions of the protein expression, the truncated proteins were expressed on a large scale and purified. Pull-down assay was performed to screen interacting partners in the total lysate and membrane-enhanced fractions of SARS-CoV susceptible cells. Afterwards, potential partners were identified by two-dimensional gel electrophoresis and mass spectrometry. Up till now, when

glutathione S-transferase (GST)-S1 was acted as bait, four potential interacting partners, but not ACE2 or other membranous protein, were identified. One of them is peroxiredoxin 1 binding to GST while others are chaperones including heat shock 70 kDa protein, GroEL and DnaJ. On the other hand, by using receptor binding domain expressed in *P. pastoris*, ACE2 was detected by western blotting, suggesting that the pull-down assay was successful to capture real interacting partner.

摘要

嚴重急性呼吸系統綜合症 (SARS) 是一種新發現的疾病，在二零零三年引發了一場全球性的災難，共造成 916 人死亡和超過 300 億美元損失。及後，該病原體被發現，並命名爲“SARS 冠狀病毒”(SARS-CoV)。在四種結構性蛋白之中，其中一種稱爲刺糖蛋白，與病毒入侵、寄主類別、組織向性及毒性有關。該糖蛋白擁有兩個含功能的次單元，稱爲 S1 和 S2，分別是負責連接受體和薄膜融合。由於病毒入侵的機制仍然未能完全理解，所以鑑定互動伙伴可能有助尋找 SARS 病發的機制，及開發有效的中和疫苗，以用作對抗 SARS-

直至現時爲止，兩個含功能的受體已被鑑定，分別是血管緊張素轉化酶 2 (ACE2) 和 CD209L。但是這些受體的組織分佈與受病毒感染的組織分佈並不相同，一些高度表達 ACE2 和 CD209L 的組織並沒有受到病毒感染。我們相信其他未被鑑定的受體或共受體仍有待發現。

在這個研究中，可譯成各種縮短刺蛋白的 DNA 片段被克隆到 pGEX-6P-1、pET-28a、pRHisMBP 和 pPICZα-A，再透過 *E. coli* 和 *P. pastoris* 表達系統表達縮短刺蛋白。經過改良蛋白表達的環境後，縮短刺蛋白被大規模表達及純化。再利用可受病毒感染的細胞的總裂解液和薄膜加強餾分，進行拉下實驗甄別互動伙伴。然後使用二維凝膠電泳和質譜儀鑑定潛在伙伴。直至目前爲止， 當 glutathione S-transferase (GST)-S1 作爲誘餌，已經發現四個可能的互動伙伴，但並不包括 ACE2 或任何細胞膜蛋白，其中一個是與 GST 連結的過氧化物酶 1 ，而另外三個伴侶蛋白分別是熱休克蛋白 70，GroEL 及 DnaJ。

另一方面，利用 *P. pastoris* 所表達的受體結合功能區，ACE2 可被西方轉漬法偵測，由此推想，拉下實驗可成功捕獲真正的互動伙伴。

CONTENTS

Materials and Methods

2. Plasmid Construction

3. Protein Expression, Purification and Analysis

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

2D	Two-dimensional
$3CL^{pro}$	Papain-like proteinase
a. a.	Amino acid
ACE2	Angiotensin-converting enzyme 2
ACN	Acetanitrate
AOXI	Alcohol oxidase I
ATP	Adenosine triphosphate
BAF	Bafilomycin A
BCoV	Bovine coronavirus
BLAST	Basic Local Alignment Search Tool
BMGY	Buffered glycerol-complex medium
BMMY	Buffered methanol-complex medium
bp	Base pair
BSA	Bovine serum albumin
C-	Carboxyl
$CaCl_2$	Calcium chloride
cDNA	Complementary deoxyribonucleic acid
CFR	Case-fatality ratio
C. I.	Confidence interval
CoV	Coronavirus
DC-SIGN	Dendritic cell-specific ICAM-grabbing non-integrin
DC-SIGNR (CD209L)	Dendritic cell-specific ICAM-grabbing non-integrin-related protein
DMEM	Dulbecco's Modified Eagle Medium
DnaJ	Heat shock protein 40
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
E	Envelope
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum

EtOH	Ethanol
FP	Fusion peptide
Grp78	Glucose-regulated protein 78
GSH	Reduced glutathione
GSSG	Oxidized glutathione
GST	Glutathione S-Transferase
guHCl	Guanidine hydrochloride
H_2O	Water
HCoV-229E	Human CoV 229E
HE	Hemagglutinin-esterase
HEL	RNA helicase
HIV	Human immunodeficiency virus
HR	Heptad Repeat
HRP	Horseradish peroxidase
HSPA5	Heat Shock 70 kDa Protein 5
IAA	Iodoacetamide
IBV	Infectious bronchitis virus
IEF	Isoelectric focusing
IPG	Immobilized pH gradient
IPTG	Iso-propylthio-β-D-galactopyranoside
kb	Kilobase pairs
KCl	Potassium chloride
kDa	Kilodalton
LB	Luria-Bertani
M	Membrane
MALDI	Matrix-Assisted Laser Desorption Ionisation
MBP	Mannose binding protein
MeOH	Methanol
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
MHV	Murine hepatitis virus
M. W.	Molecular weights
N	Nucleocapsid

N-	Amino
NaCl	Sodium chloride
NH_4Cl	Ammonium chloride
NH_4HCO_3	Ammonium bicarbonate
$NiSO_4$	Nickel sulphate
OD	Optical density
ORFs	Open reading frames
PBS	Phosphate buffer saline
PBST	Phosphate buffer saline - polyoxyethylenesorbitan monolaurate
PCR	Polymerase chain reaction
PEDV	Porcine epidemic diarrhea virus
PEG	Polyethylene glycol
PMFs	Peptide mass fingerprints
PMSF	Phenylmethylsulfonyl fluoride
PNGase F	Peptide-N-glycanase F
POL	RNA-dependent RNA polymerase
Prx1	Peroxiredoxin 1
r.p.m.	Revolutions per minute
RbCl	Rubidium chloride
RBD	Receptor-binding domain
RNA	Ribonucleic acid
S	Spike
SARS	Severe Acute Respiratory Syndrome
SARS-CoV	Severe Acute Respiratory Syndrome-associated Coronavirus
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
S/N	Signal-to-noise
TAE	Tris(hydroxymethyl)aminomethane-acetate-ethylenediaminetetraacetic acid
TFA	Trifluoroacetic acid
TGEV	Transmissible gastroenteritis virus
TM	Transmembrane

TOF	Time of flight
Tris-HCl	Tris(hydroxymethyl)aminomethane-hydrochloride
WHO	World Health Organization
YPD	Yeast extract peptone dextrose
YPDS	Yeast extract peptone dextrose sorbitol

ACKNOWLEDGEMENTS

As a beginner worked on proteomic researches in MMW610, I faced a lot of problems. However, I was fortunate that many people gave helping hands hence I could complete this dissertation. First of all, I would like to express my deepest thanks to my supervisor, Prof. Stephen K. W. Tsui, for providing me an atmosphere of research freedom, valuable advices and encouragement. I would also like to express my grateful gratitude to Prof. Shannon W. N. Au, Prof. K. B. Wong and Prof. Mary M. Y. Waye, who not only give me helpful insights but also generously share a lot of materials and apparatus.

I would like to express my special acknowledgement to Miss Meiji Ma, who unconditionally shares a lot of precious ideas and experiences as well as provides a working environment with endless joy. Also, I am grateful to Mr. Thomas Au, Miss Karman Leung and Mrs. Yeung-Yeo Yuk Yin for their tremendous assistance in solving some problems in my study, especially molecular cloning and mass spectrometry. And I would like to express my sincerely thank to my friends and labmates in MMW610, BMSB316 and MMW612A, particularly Miss Cecilia Chan, Mr. Yeung Tze Lun, Miss Samantha Lau, Miss Sophie Chan, Miss Frances Ho, Miss Crystal Cheung, Miss Cherry Chau and Mr. Lee Kai Woo. Their intellectual and spiritual support is supreme. Moreover, I express my extraordinary gratitude to Miss Mo Hung Ting for her gracious concerns and gifts. Besides, I am grateful to my best friends for sharing happiness and sadness. Their kindnesses cannot be replaced.

Last but not least, I dedicate this thesis to my grandfather, my parents and Miss Tracy Tsang to thank for their support, encouragement and patience. They made every enterprise to devote their time and effort to my life.

Background

1.1 SARS

1.1.1 Outbreak and Influence

On 16 November 2002, the first case of severe atypical pneumonia of unknown cause in the world was reported for the Guangdong Province, Southern China (Zhao *et al.*, 2003b). After three months of the first reported case, outbreak of this atypical pneumonia had been occurred in Guangdong Province that the patient number was raised to 305 [World Health Organization (WHO), 2003a; WHO, 2003b; Zhong *et al.*, 2003]. Meanwhile, cases from other countries and regions including Hong Kong, Vietnam, Canada and Singapore were reported (Centers for Disease Control and Prevention, 2003; Lee *et al.*, 2003; Poutanen *et al.*, 2003; Reilley *et al.*, 2003). This novel atypical pneumonia was later named Severe Acute Respiratory Syndrome (SARS) by WHO. SARS patients were treated by various drugs, for instance, interferons, ribavirin and human immunodeficiency virus (HIV) protease inhibitors, that were however not highly effective (Cinatl *et al.*, 2004). Up to 5 July 2003, absence of new reported case represented SARS outbreak termination. Within eight months, cumulatively 8,422 patients from 32 countries or regions were infected and 916 of them were dead (Table 1.1) (WHO, 2003c).

The average case-fatality ratio (CFR), which measures the patients who die from a disease and reflects virulence, in the world was 11 %. Mainland China had the highest SARS patient number, which was 5,327, while the region of the highest attack rate was Hong Kong, which 25.6 inhabitants were infected in every 100,000 inhabitants (Figure 1.1) (WHO, 2003c). Besides lives loss, SARS induced loss of approximately 30 billion United States Dollars in the worldwide (Lemonick *et al.*,

2003).

Table 1.1 – The numbers, ages and CFRs of SARS patients in selected areas. Female and male cumulative numbers of cases from Mainland China were excluded (Adapted and modified from http://www.who.int/csr/sars/country/en/country2003_08_15.pdf).

Areas	Cumulative number of cases			Median age (range)	CFR (%)
	Female	Male	Total		
China	/	/	5327	Pending	7
Hong Kong	977	778	1755	40 (0-100)	17
Taiwan	349	319	665	46 (2-79)	27
Canada	151	100	251	49 (1-98)	17
Singapore	161	77	238	35 (1-90)	14
Vietnam	39	24	63	43 (20-76)	8
United States	16	17	33	36 (0-83)	0
Philippines	8	6	14	41 (29-73)	14
Germany	4	5	9	44 (4-73)	0
Mongolia	8	1	9	32 (17-63)	0
Thailand	5	4	9	42 (2-79)	22
France	1	6	7	49 (26-61)	14
Australia	4	2	6	15 (1-45)	0
Malaysia	1	4	5	30 (26-84)	40
Italy	1	3	4	30.5 (25-54)	0
United Kingdom	2	2	4	59 (28-74)	0
India	0	3	3	25 (25-30)	0
Korea	0	3	3	40 (20-80)	0
Sweden	1	2	3	33	0
Indonesia	0	2	2	56 (47-65)	0
Brazil	1	0	1	4	0
Macao	0	1	1	28	0
Colombia	1	0	1	28	0
Finland	0	1	1	24	0
Kuwait	1	0	1	50	0
New Zealand	1	0	1	67	0
Ireland	0	1	1	56	0
Romania	0	1	1	52	0
Russian	0	1	1	25	0
South Africa	0	1	1	62	100
Spain	0	1	1	33	0
Switzerland	0	1	1	35	0
Total	1732	1366	8422	/	11

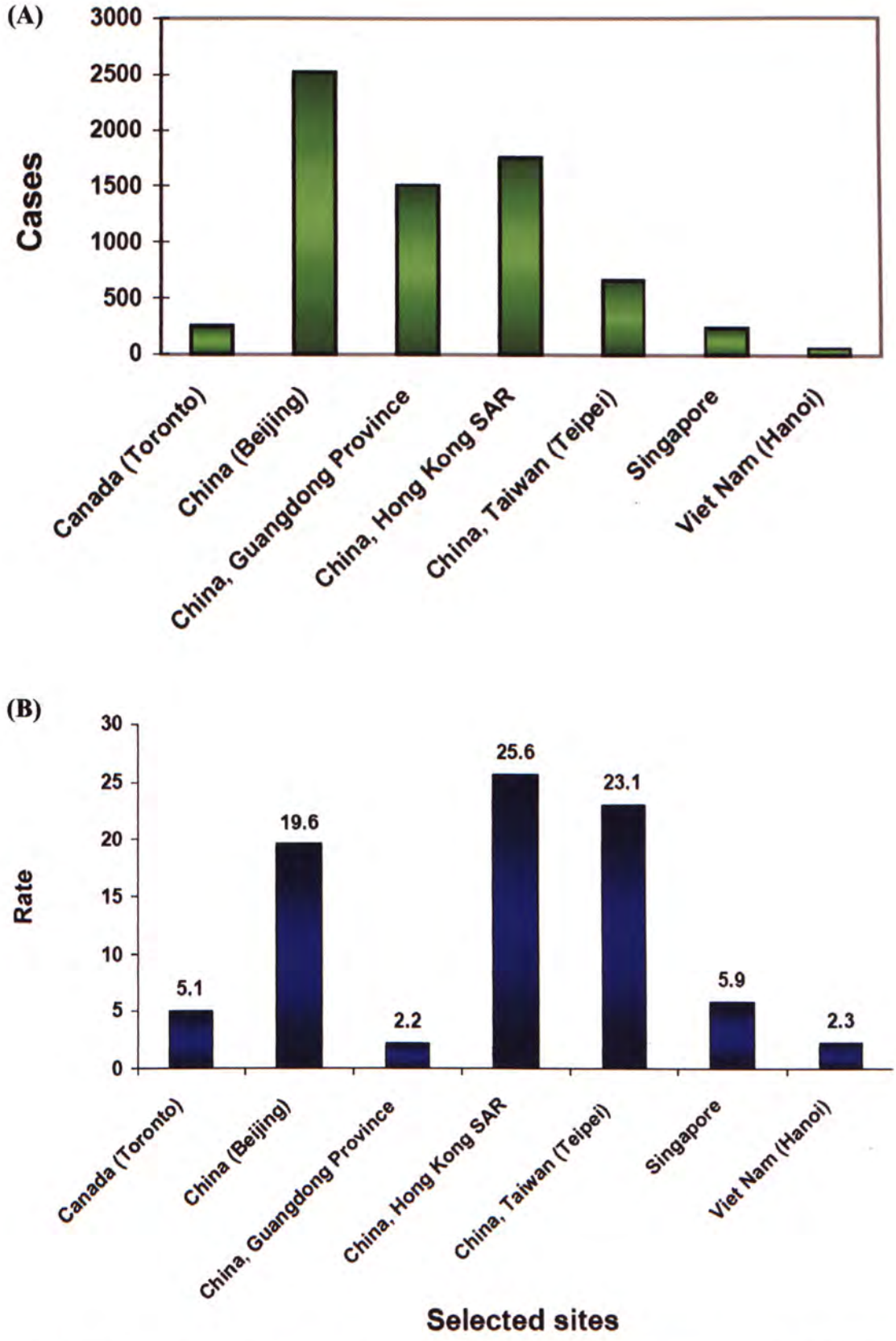


Figure 1.1 – The numbers and attack rates of SARS patients in selected sites. **(A)** The SARS patient number and **(B)** attack rates in selected sites were shown. The highest SARS patient number was in Beijing while the highest attack rate was in Hong Kong (Adapted and modified from http://www.who.int/csr/sars/en/WHO consensus.pdf).

1.1.2 Clinical Features

Cases of SARS patients in all ages had been reported. A possible reason for the slightly higher number of female patients was the higher chance of exposure among nurses (Donnelly *et al.*, 2003). The incubation period was between two and ten days while its average was five days (Chan *et al.*, 2004b; Chan-Yeung *et al.*, 2003; Tsang *et al.*, 2003; Wu *et al.*, 2003; Zhao *et al.*, 2003a). Fever presented in more than 99 % of the patients was the most common symptom. Cough, malaise and chill presented in 65.5 %, 58.8 % and 51.5 % of the patients respectively were also usual. Others such as myalgia, dyspnea and tachycardia were detected in some patients (Table 1.2) (Lee *et al.*, 2003; Peiris *et al.*, 2003a; Peiris *et al.*, 2003b; Zhao *et al.*, 2003a).

Table 1.2 – Sexes, clinical features and physical signs of selected SARS patients (Adapted and modified from Peiris *et al.*, 2003a).

	No. of patient with symptom	No. of total patient	Percentage (%)
Sex of patients			
Male	343	752	45.6
Female	409		54.4
Clinical features			
Fever	751	752	99.9
Cough	460	702	65.5
Malaise	317	539	58.8
Chill or rigors	377	732	51.5
Myalgia	365	752	48.5
Dyspnea	282	614	45.9
Dizziness	163	597	27.3
Nausea or vomiting	8	30	26.7
Chest pain or pleurisy	47	210	22.4
Diarrhea	130	647	20.1
Anorexia	37	188	19.7
Headache	292	752	18.8
Sore throat	91	552	16.5
Rhinorrhoea	50	362	13.8
Physical signs			
Tachycardia	71	154	46.1
Tachypnea	60	154	39.0
Chest rales	56	204	27.5

1.2 SARS-CoV

1.2.1 Genomic Organization

The pathogen of SARS is a novel coronavirus (CoV) named SARS-associated CoV (SARS-CoV), which was isolated from SARS patients (Drosten *et al.*, 2003; Ksiazek *et al.*, 2003; Peiris *et al.*, 2003b). Experimental data also demonstrated that SARS-CoV is the cause of SARS (Fouchier *et al.*, 2003; Kuiken *et al.*, 2003). The viral genetic material of 30 kilobase pairs (kb) is single-stranded plus sense ribonucleic acid (RNA) possessing 15 putative open reading frames (ORFs) (Table 1.3) (Ksiazek *et al.*, 2003; Marra *et al.*, 2003; Rota *et al.*, 2003; Thiel *et al.*, 2003; Zeng *et al.*, 2003).

Among these ORFs, ORF1a and ORF1b comprising approximately two-thirds of the genome are translated to two polyproteins that are divided into a number of replicases by proteolytic processing. These replicases include RNA-dependent RNA polymerase (POL), RNA helicase (HEL) and papain-like proteinase ($3CL^{pro}$) (Gao *et al.*, 2003a; Snijder *et al.*, 2003; Thiel *et al.*, 2003). Their functions are replication of viral RNA and proteins, but the detailed replication mechanisms are not fully understood.

Another four ORFs encoding structural proteins, spike (S), envelop (E), membrane (M) and nucleocapsid (N) proteins can be found in the SARS-CoV genome. The major role of N protein is the association with viral RNA, leading to helical N formation (Chang *et al.*, 2006; Hsieh *et al.*, 2005; Huang *et al.*, 2004). Abundant M protein binding to S, E and N proteins comprises intracellular membrane of SARS-CoV (Fang *et al.*, 2005; He *et al.*, 2004; Lee *et al.*, 2005; Luo *et al.*, 2006). Besides M protein, E protein is another essential component constructing the viral membrane and binding to S protein (Ho *et al.*, 2004b; Wu *et al.*,

2003). Finally, S protein, a transmembrane (TM) protein located at the viral membrane, can bind to cellular receptor and mediate viral entry (Giroglou *et al.*, 2004; He *et al.*, 2006; Sainz *et al.*, 2006; Yeung *et al.*, 2006).

In addition, nine putative accessory proteins of more than 50 amino acid (a. a.) residues are predicted to be translated though their presence and functions are still under investigation. Lastly, s2m motif, a conserved RNA sequence found in other CoVs, was detected at the 3' regions of the genomes of all SARS-CoV strains (Robertson *et al.*, 2005). The gene encoding hemagglutinin-esterase (HE), which is commonly found in other CoVs, is absent in the SARS-CoV genome, suggesting that this protein is unrelated to SARS-CoV replication (Marra *et al.*, 2003; Rota *et al.*, 2003).

Table 1.3 – Features of SARS-CoV genome sequences and subgenomic transcripts (Adapted and modified from Chow *et al.*, 2003).

g/sg mRNA	ORF				Start-End	No. of a.a.	No. of Bases	Frame
	Thiel *et al.*	Zeng *et al.*	Marra *et al.*	Rota *et al.*				
mRNA 1	ORF 1a	ORF 1a	ORF 1a	ORF 1a	265-13,398	4,382	13,149	+1
mRNA 1	ORF 1b	ORF 1b	ORF 1b	ORF 1b	13,398-21,485	2,628	7,887	+3
mRNA 2	S protein	S protein	S protein	S protein	21,492-25,259	1,255	3,768	+3
mRNA 3	ORF 3a	X1	ORF 3	X1	25,268-26,092	274	825	+2
mRNA 3	ORF 3b	N/R	ORF 4	X2	25,689-26,153	154	465	+3
mRNA 4	E protein	N/R	E protein	E protein	26,117-26,347	76	231	+2
mRNA 5	M protein	M protein	M protein	M protein	26,398-27,063	221	666	+1
mRNA 6	ORF 6	N/R	ORF 7	X3	27,074-27,265	63	192	+2
mRNA 7	ORF 7a	X2	ORF 8	X4	27,273-27,641	122	369	+3
mRNA 7	ORF 7b	N/R	ORF 9	N/R	27,638-27,772	44	135	+2
mRNA 8	ORF 8a	X3	ORF 10	N/R	27,779-27,898	39	120	+2
mRNA 8	ORF 8b	N/R	ORF 11	X5	27,864-28,118	84	255	+3
mRNA 9	N protein	N protein	N protein	N protein	28,120-29,388	422	1,269	+1
mRNA 9	ORF 9b	N/R	ORF 13	N/R	28,130-28,426	98	297	+2

1.2.2 Morphology

SARS-CoV particles were observed in some SARS patient tissues and SARS-CoV-cultivating cells under electron microscopes. The virion is spherical and about 60 – 140 nm in diameter. N comprising RNA genome and N protein is located in the centre, while the inner core surrounded by outer lipid envelop is made of S, M and E proteins. On the virion surface, S protein appearing as complex projections with 20 – 40 nm in length is similar to that of other CoVs (Figure 1.2 and 1.3) (Ksiazek *et al.*, 2003; Tse *et al.*, 2004; Zhang *et al.*, 2003).

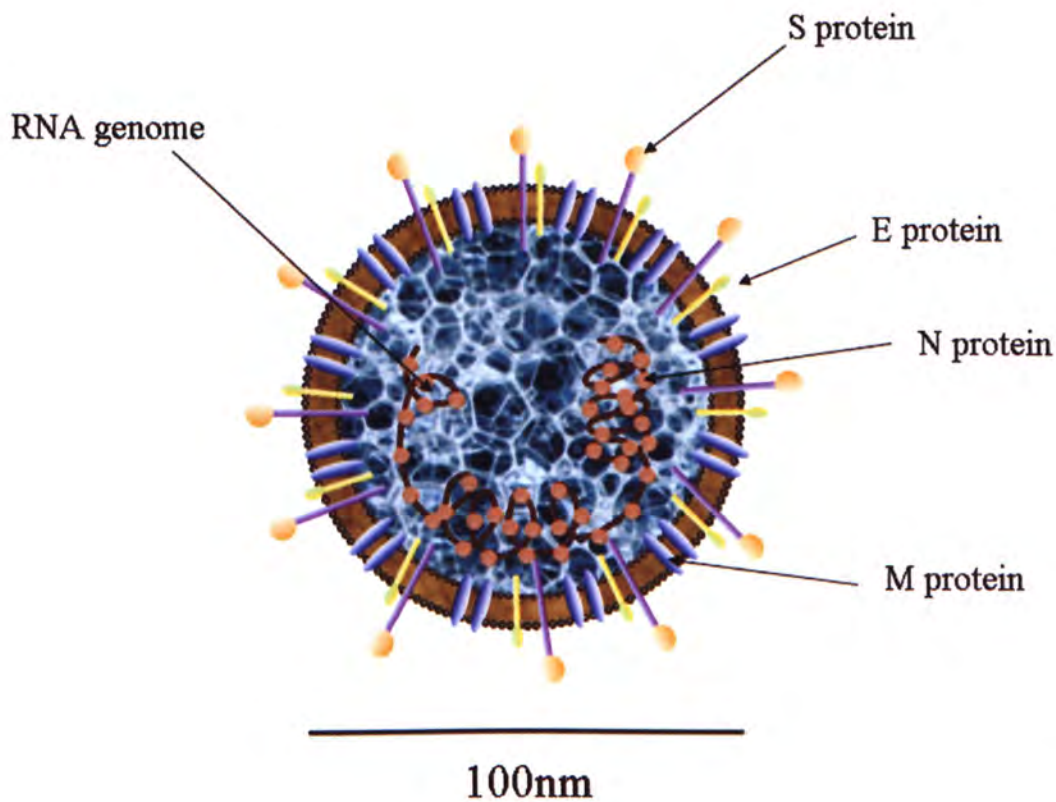
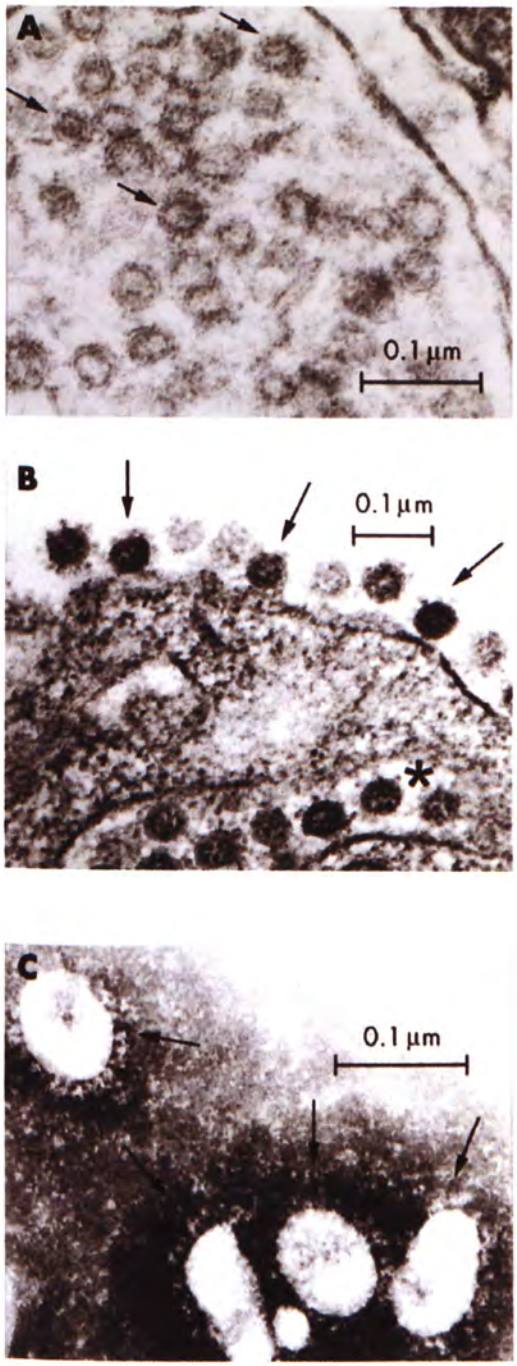


Figure 1.2 – Schematic diagram of SARS-CoV. SARS-CoV appearing as a 100 nm particle with a corona-like structure is composed of S, E, M and N proteins (Adapted and modified from http://www.sarsreference.com/sarsreference.pdf).



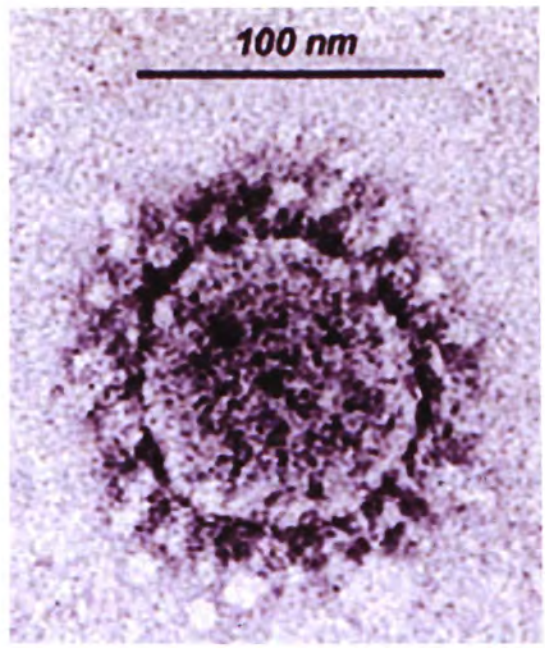


Figure 1.3 – Morphology of SARS-CoV under electronic microscopy. **(A)** SARS-CoV particles were found in dilated cytoplasmic vesicles in lung cells of SARS patients. The arrows indicate the particles of the better preserved structures. **(B)** The asterisk arrows indicate similar viral particles found in endoplasmic reticulum (ER) and on the surface of the cytoplasmic membrane of infected Vero cell respectively. **(C)** The particles were found in the supernatant of Vero cell culture. **(D)** SARS-CoV with an internal helical N-like structure was surrounded by club-shaped complex projections. All bars represent 100 nm (Adapted and modified from Ksiazek *et al.*, 2003 and Tse *et al.*, 2004).

1.2.3 Phylogenetic Analysis

Before SARS outbreak, more than ten identified CoVs were divided into three subgroups according to their genome sequences (Horzinek *et al.*, 1999). Protein sequence alignment of replicases and structural proteins of SARS-CoV and other CoVs showed their identities are approximately between 20 % and 60 %, suggesting that they were diverse (Table 1.4) (Rota *et al.*, 2003). From the phylogenetic trees constructed by nucleic acid and protein sequences of CoVs, SARS-CoV was located between group II and group III of the CoV family (Figure 1.4) (Eickmann *et al.*, 2003; Ksiazek *et al.*, 2003; Marra *et al.*, 2003; Rota *et al.*, 2003; Spiga *et al.*, 2003). The location further indicated SARS-CoV is not belonged to any subgroup and possible to be the first member of a new group of CoV. Nevertheless, some different phylogeny analyses demonstrated contradictory – SARS-CoV belongs to a sub-group of group II rather than a new group (Gibbs *et al.*, 2004; Gorbalenya *et al.*, 2004; Kim *et al.*, 2006).

Table 1.4 – Pairwise identities of CoV proteins. Sequences and lengths of SARS-CoV proteins were compared with that of selected CoVs in the three groups including human CoV 229E (HCoV-229E), porcine epidemic diarrhea virus (PEDV), transmissible gastroenteritis virus (TGEV), bovine CoV (BCoV), murine hepatitis virus (MHV) and infectious bronchitis virus (IBV) (Adapted and modified from Rota *et al.*, 2003).

Group	Virus	Pairwise Amino Acid Identity (Percent)						
		3CLPRO	POL	HEL	S	E	M	N
G1	HCoV-229E	40.1	58.8	59.7	23.9	22.7	28.8	23.0
	PEDV	44.4	59.5	61.7	21.7	17.6	31.8	22.6
	TGEV	44.0	59.4	61.2	20.6	22.4	30.0	25.6
G2	BCoV	48.8	66.3	68.3	27.1	20.0	39.7	31.9
	MHV	49.2	66.5	67.3	26.5	21.1	39.0	33.0
G3	IBV	41.3	62.5	58.6	21.8	18.4	27.2	24.0

	Virus	Predicted Protein Length (aa)						
	SARS-CoV	306	932	601	1255	76	221	422
	CoV Range	302-307	923-940	506-600	1173-1452	76-108	225-262	377-454

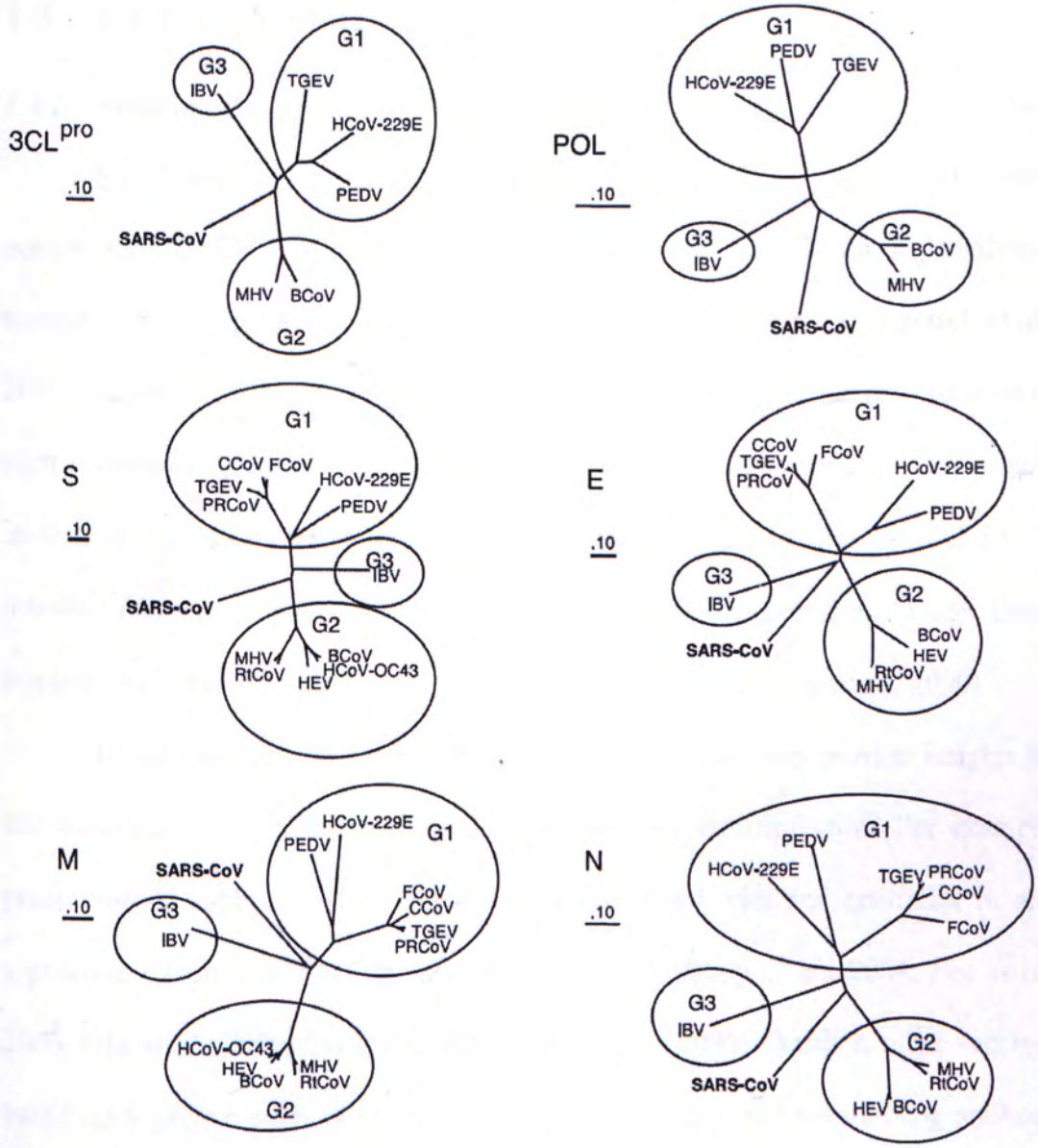


Figure 1.4 – Phylogenetic analysis of CoV proteins. Sequence alignments and neighbor-joining trees were generated by comparing $3CL^{PRO}$, POL, S, E, M and N proteins between SARS-CoV and six of CoVs in the three sub-groups. All SARS-CoV proteins were distant from that of other CoVs (Adapted and modified from Rota *et al.*, 2003).

1.3 S Glycoprotein

1.3.1 Functional Roles

S glycoprotein, one of the structural proteins of SARS-CoV, is a TM protein passing through the viral membrane. It is approximately 20 – 40 nm in length and surrounds the periphery of SARS-CoV surface (Figure 1.2 and 1.3) (Ksiazek *et al.*, 2003; Tse *et al.*, 2004; Zhang *et al.*, 2003). S plays an important role on mediation of viral invasion via receptor binding and membrane fusion. The presence of S protein as well as interaction between S protein and receptor are essential for SARS-CoV infection (Simmons *et al.*, 2004). Hence, this protein can determine host range, tissue tropism, and virulence of SARS-CoV (Gallagher *et al.*, 2001; Kuo *et al.*, 2000).

Investigation of the protein mediating viral invasion may provide insights for the identification of effective methods to prevent the viral infection. For example, production of antibody, which can specifically interact with and neutralize S, was a possible effective and safety vaccines for therapy (Choy *et al.*, 2004; Lee *et al.*, 2006; Qin *et al.*, 2006; Sui *et al.*, 2004; Zhou *et al.*, 2006). Besides, other vaccines based on S protein such as DNA vaccine inducing T cell and neutralizing antibody responses, as well as RNA interference inhibiting S expression were developed (Gao *et al.*, 2003b; Li *et al.*, 2005b; Wang *et al.*, 2005; Yang *et al.*, 2004b; Zeng *et al.*, 2004; Zhang *et al.*, 2003; Zhao *et al.*, 2004). Other than vaccination, development of antibodies recognizing S may help in diagnosis of SARS patients in early stage (Chang *et al.*, 2004; Ohnishi *et al.*, 2005).

1.3.2 Structure and Functional Domains

S glycoprotein of SARS-CoV has 1,255 a. a. residues and its theoretical molecular weight is 139 kilodalton (kDa) (Ksiazek *et al.*, 2003; Marra *et al.*, 2003; Rota *et al.*, 2003). However, the actual molecular weight is 180 kDa due to the presence of at least four glycosylation sites (Krokhin *et al.*, 2003; Ying *et al.*, 2003). Alignment results using various CoV S protein sequences revealed that S protein of SARS-CoV is only approximately 25 % identical (Table 1.4) (Rota *et al.*, 2003). Even though the identity is low; its structure can still be predicted to be a type I TM protein and divided into two functional domains, S1 and S2, located extracellularly near amino (N-) and carboxyl (C-) terminals respectively (Figure 1.5). Their functions are binding cellular receptor and mediating fusion between viral and cellular membranes respectively. By comparing with other CoVs, S1 of SARS-CoV is less conserved than S2 (Liu *et al.*, 2004).

At the N-terminal of S1 domain, a signal peptide, which has presumably been removed during transport through ER, of approximately 12 a. a. residues was identified (Ho *et al.*, 2004a). Amino acid residues 17 – 276 in form of dimer are involved in membrane fusion but not receptor-binding (Xiao *et al.*, 2004). Also, a. a. residues 318 – 510 were identified as receptor-binding domain (RBD) interacting with angiotensin-converting enzyme 2 (ACE2) (More details can be found in section 1.3.3, p.15) (Li *et al.*, 2003; Wong *et al.*, 2004).

Between S1 and S2, a proteolytic cleavage site commonly found in most of Group II and Group III CoVs is absent in SARS-CoV (Rota *et al.*, 2003). No evidence demonstrated that SARS-CoV S protein is cleaved in any stage as well (Moore *et al.*, 2004; Simmons *et al.*, 2004; Song *et al.*, 2004; Xiao *et al.*, 2003). Absence of the cleavage site causes the exact boundary between S1 and S2 undefined,

though a putative site at a. a. residue of 672 is predicted (Xiao *et al.*, 2003). The cleavage of SARS-CoV is unnecessary for viral infection but enhances membrane fusion (Bosch *et al.*, 2004; Follis *et al.*, 2006; Simmons *et al.*, 2004; Song *et al.*, 2004; Xiao *et al.*, 2003).

Presence of fusion peptide (FP), a short peptide upstream of HR1, in SARS-CoV S2 domain is predicted as other CoVs. This hydrophobic peptide is rich in alanine, glycine and phenylalanine residues, and possesses a proline residue. During viral entry, FP exposes and associates with host cell membrane leading to membrane fusion. The predicted FP location is approximately a. a. residues 858 – 886 (Bosch *et al.* 2004; Ingallinella *et al.*, 2004; Petit *et al.*; Supekar *et al.* 2004). Nonetheless, a. a. residues 770 – 788 inserting into the lipid membrane were demonstrated (Sainz *et al.*, 2005), so the exact location of FP in SARS-CoV remains to be determined.

Amino acid residues 896 – 972 and 1142 – 1188 are Heptad Repeat 1 (HR1) and 2 (HR2) respectively, which play important roles for membrane spinning (More details can be found in section 1.3.4, p.17) (Bosch *et al.*, 2004; Xu *et al.*, 2004b; Zhu *et al.*, 2004). They are helical and themselves assemble into a coiled coil structure (Ingallinella *et al.*, 2004; Tripet *et al.*, 2003). The HR1 connects FP and HR2 while the HR2 association in native pre-fusogenic stage leads to trimerization of S protein (More details can be found in section 1.3.4, p.17) (Hakansson-McReynolds *et al.*, 2006).

At S2 C-terminal, regions located inside the viral membrane and inner core compartment are named TM domain and cytoplasmic tail respectively. TM is critical for viral infection and stabilization of S protein trimerization (Broer *et al.*, 2006; Song *et al.*, 2004; Xiao *et al.*, 2004).

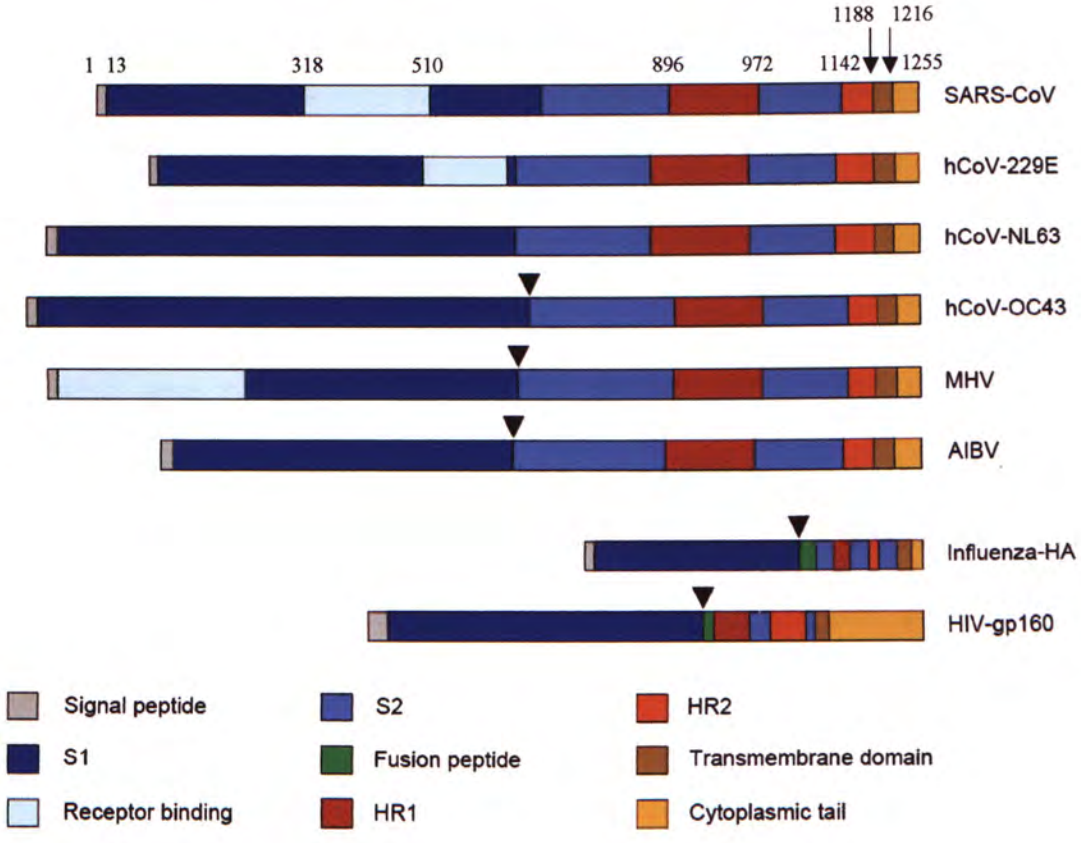


Figure 1.5 – Structural features of selected CoV S proteins (Adapted and modified from Hofmann *et al.*, 2004c).

1.3.3 Interacting Partners

ACE2, a 120 kDa type I (C-terminus intracellular) integral membrane protein mainly expressed in heart, kidney and testes, lung and intestine, was identified as a functional receptor of SARS-CoV (Douglas *et al.* 2004; Hofmann *et al.*, 2004a; Li *et al.*, 2003; Tikellis *et al.*, 2003; Wang *et al.*, 2004). SARS-CoV infection and replication were detected in ACE2-transfected 293T cells, which are originally SARS-CoV non-susceptible cells (Figure 1.6). Also, viral infection on VeroE6 cells, that were SARS-CoV susceptible cells, was blocked by anti-ACE2 antibody (Figure 1.7). ACE2 is a zinc metallopeptidase cleaving one a. a. from the C-terminal of substrates and is important for regulation of cardiac function (Averill *et al.*, 2003; Crackower *et al.*, 2002; Donoghue *et al.*, 2000; Tipnis *et al.* 2000; Zisman *et al.*, 2003). Its peptidase domain is proved to be assembled with a. a. residues of 318 – 510 of SARS-CoV S protein (Babcock *et al.*, 2004; Li *et al.*, 2005a; Towler *et al.*, 2004; Wong *et al.*, 2004; Xiao *et al.*, 2003).

On the other hand, interactions between S glycoprotein and two kinds of TM C-type (calcium dependent) lectin, dendritic cell-specific ICAM-grabbing non-integrin (DC-SIGN) and its homologues DC-SIGN-related protein (DC-SIGNR / CD209L), were demonstrated (Marzi *et al.*, 2004; Jeffers *et al.*, 2004). DC-SIGN is expressed on DC surface and some types of macrophages while DC-SIGNR is located on sinusoidal endothelial cells and placental macrophages (Bashirova *et al.*, 2001; Geijtenbeek *et al.*, 2000; Soilleux *et al.*, 2001). The major physiological role of DC-SIGN is to interact with conserved molecules of some pathogens and then initiates immune responses against the pathogens (van Kooyk *et al.*, 2003). They are however served as attachment factors, enhancing the interaction between SARS-CoV S protein and other interacting partners, rather than functional receptors (Marzi *et al.*,

2004). In HIV-infected cells, they act as attachment factors associating with a range of viral proteins including the envelope protein (Alvarez *et al.*, 2002; Gardner *et al.*, 2003; Geijtenbeek *et al.*, 2000; Klimstra *et al.*, 2003; Lozach *et al.*, 2003).

Figure 1.6 – SARS-CoV infection in SARS-CoV non-susceptible ACE2-transfected cell line. The presence of viral RNA in 293T cells medium in various dilutions was detected by RT-PCR. RNA was detected in 293T cells with ACE2-transfection but not mock-transfection control (Adapted and modified from Li *et al.*, 2003).

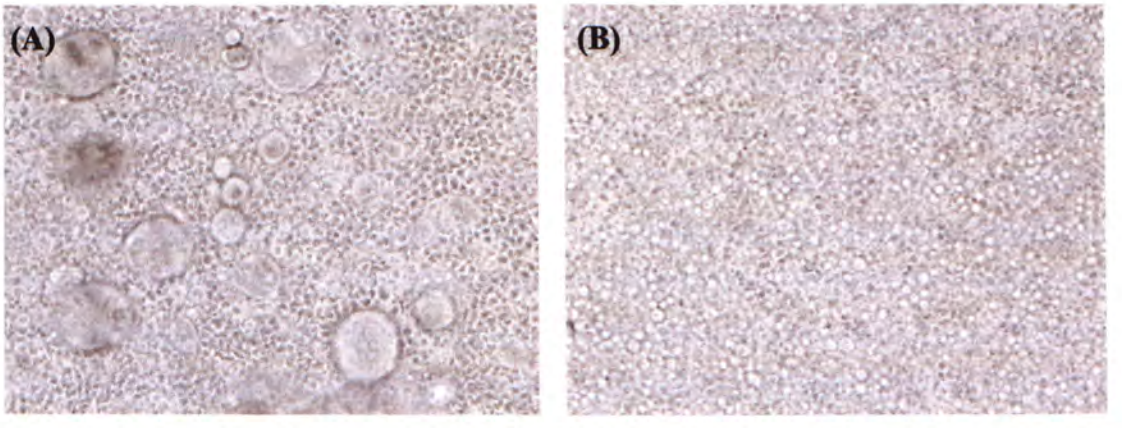


Figure 1.7 – Blockage of SARS-CoV infection by anti-ACE2 antibody. **(A)** Syncytia formation was determined between S protein- and ACE2-expressing 293T cells. **(B)** The syncytia formation was inhibited by anti-ACE2 antibody (Adapted and modified from Li *et al.*, 2003).

1.3.4 Viral Entry Mechanism

S glycoprotein binding to functional receptor on the host cell surface is the initial step of viral entry. Afterwards, SARS-CoV is predicted to enter host cells by two pathways as other enveloped viruses (Figure 1.8) (Eckert *et al.*, 2001; Hofmann *et al.*, 2004a; Smith *et al.*, 2004; Yeung *et al.*, 2006). In the pH-dependent pathway, receptor binding causes internalization of virion into endosomes by endocytosis. In the presence of hydrogen ion pump on the membrane, acidification of endosomal compartment induces conformational change in the S protein. With regards to the pH-independent pathway, receptor-binding directly triggers change in S protein conformation.

In some cases, viral infection can be inhibited by ammonium chloride (NH_4Cl), which neutralizes the acidic environment of endosome (Hofmann *et al.*, 2004b; Nie *et al.*, 2004; Simmons *et al.*, 2004; Yang *et al.*, 2004a), suggesting that SARS-CoV pathogenesis is a pH-dependent endocytotic process. Viral entry via membrane fusion at the cell surface was also observed by microscopy (Zhang *et al.*, 2004). However, the exact viral entry pathway of SARS-CoV remains to be investigated.

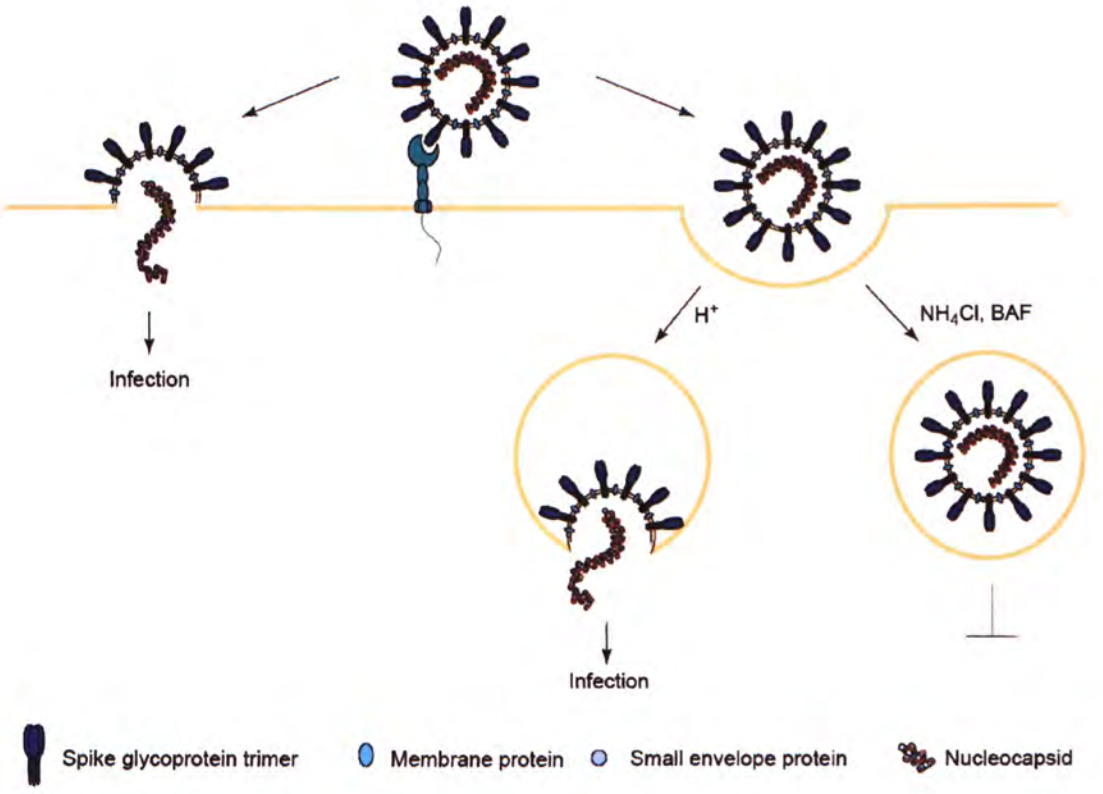


Figure 1.8 – Schematic diagram of two putative pathways triggering conformational change in SARS-CoV S protein. The left is the pH-independent pathway that receptor-binding directly triggers change in S protein conformation. Viral and plasma membranes are fused and hence N is released to cytoplasm. The right is the pH-dependent pathway that receptor-binding induces virion is up taken into the endosome. Acidification of endosomal compartment induces the conformational alteration leading to membrane fusion. The acidification can be inhibited by NH_4Cl or bafilomycin A (BAF) and thus blocks the viral entry (Adapted and modified from Hofmann *et al.*, 2004c).

After S1 dissociation, FP covered in native pre-fusogenic stage exposes and inserts to host cell membrane surface (Figure 1.9). Next, in fusion active state, HR2 folds back onto HR1 leading to a six-helix bundle structure (trimer of dimers) formation. HR1 associates into a central trimeric coiled coil structure and HR2 are in anti-parallel orientation, thereby bringing the host cell membrane and SARS-CoV TM domain into close contact, and hence promoting membrane fusion (Duquerroy *et al.*, 2005). Afterwards, viral N can be transferred to host cell cytoplasm for viral replication (Hofmann *et al.*, 2004c).

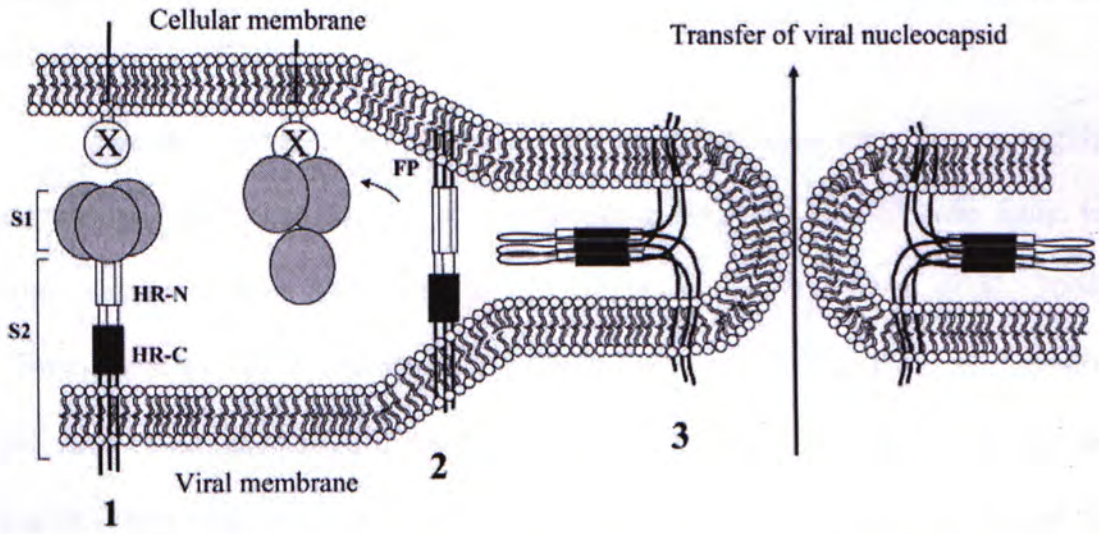


Figure 1.9 – Schematic diagram of the three states of SARS-CoV S protein during viral entry. (1) S protein in native pre-fusogenic stage does not bind to any receptor. **(2)** In intermediate state, after binding, S1 domain dissociates to expose FP which inserts to host cell membrane. **(3)** The S2 domain in fusion active state collapses to draw the viral and host cell membranes together leading to membrane fusion and hence transfer of viral N (Adapted and modified from Tripet *et al.*, 2004).

1.4 Aim of Study

1.4.1 Mismatch of SAR S-CoV Tissue Tropism and Tissue Distribution of ACE2

Tissues and cell lines expressing functional receptors are theoretically susceptible to SARS-CoV, and vice versa. SARS-CoV replication leads to the presence of viral RNA, proteins and virions in the infected tissues of SARS patients. The comparison results indicated the general pattern of ACE2 expression roughly agrees with SARS-CoV tropism. For instance, lung and intestine expressing ACE2 were the major infected organs where viral RNA, proteins and particles were identified (Table 1.5) (Chow *et al.*, 2004; Ding *et al.*, 2004; Hamming *et al.*, 2004; Harmer *et al.*, 2002; Leung *et al.*, 2003; To *et al.*, 2004a; To *et al.*, 2004b; Tse *et al.*, 2003).

On the contrary, discrepancies were appeared since some tissues highly expressing ACE2 were SARS-CoV insusceptible, while SARS-CoV was found in some tissues without ACE2 (Table 1.5) (Chau *et al.*, 2004; Ding *et al.*, 2004; Hamming *et al.*, 2004; Harmer *et al.*, 2002; To *et al.*, 2004a; To *et al.*, 2004b). No virus could be detected in heart as well as endothelium of most organs that are highly expressing ACE2, whereas viral RNA and proteins were presented in hepatocytes without ACE2.

Similar results were determined during studying susceptibilities of various cell lines. SARS-CoV was replicated in VeroE6 cells expressing ACE2 but not in HepG2, HeLa, MDCK and LLC-MK2 without ACE2 (Chan *et al.*, 2003; Ng *et al.*, 2003a; Ng *et al.*, 2003b). Nonetheless, a few human colorectal adenocarcinoma cell lines expressing ACE2, including HCT-116, DLD-1, HT-29, SW-480, LS-180 and SW-620, were non-permissive to viral infection (Table 1.5) (Chan *et al.*, 2004a).

Table 1.5 – Comparison between selected SARS-CoV susceptible tissues and tissue distribution of ACE2 (Chau *et al.*, 2004; Chow *et al.*, 2004; Ding *et al.*, 2004; Hamming *et al.*, 2004; Harmer *et al.*, 2002; Leung *et al.*, 2003; To *et al.*, 2004a; To *et al.*, 2004b; Tse *et al.*, 2004).

	SARS-CoV susceptible cells	SARS-CoV insusceptible cells
Tissues with ACE2	**Lung** - Alveolar epithelia - Pneumocytes - Tracheal / bronchial serous gland epithelia **Intestine** - Surface enterocytes - Epithelial cells - Gastric parietal cells **Kidney** - Distal convoluted tubules - Epithelial cells of distal convoluted renal tubules - Cerebral neurons - Adrenal cortical cells - Medullary cells **Parathyroid and pituitary** - Acidophilic cells **Sweat gland**	**Kidney** - Glomerular viscerals - Parietal epithelial cells - Proximal tubules **Testis** - Seminiferous tubules **Heart** **Thyroid** **Oesophagus** **Spleen** **Bone marrow** **Ovary** **Uterus** **Aorta** **Cerebellum** **Muscle** *(Endothelial cells in most of organs)*
Tissues without ACE2	**Liver** - Hepatocytes - Kupffer cells - Sinusoidal endothelia	

1.4.2 Presence of Other Interacting Partner(s)

Mismatch between SARS-CoV tissue tropism and tissue distribution of ACE2 indicates the *presence* of ACE2 alone is insufficient for viral invasion. Other attachment factor(s) or co-receptor(s) enhancing the viral invasion ability may hence be present. In susceptible cells without ACE2, other functional receptor(s) involved in viral infection is possible to be existed as well. Collectively, we believe that other unidentified interacting partner(s) involved in SARS pathogenesis remains to be discovered.

1.4.3 Significance of the Study

Identification of interacting partners involved in SARS pathogenesis may be beneficial to investigate SARS-CoV entry mechanism, which is still remaining unclear. The results of this project may also provide insights into the investigation of SARS pathogenesis. Since RBD of S presumably is the most effective target epitope, construction of antibody binding to new RBD may block the interaction between virus and receptor. Last but not least, investigation of interactions between S and its interacting partners in other organisms may provide information of origins of SARS-CoV in order to have a better control of the future SARS outbreak.

Plasmid Construction

2.1 Fragment Design

2.1.1 Functional Domain Analysis

S1 and S2, the two functional domains of the SARS-CoV S glycoprotein responsible for binding cellular receptor and mediating membrane fusion respectively, were selected for expression (Figure 2.1). Since the signal peptide in S1 and the TM domain in S2 were hydrophobic, truncated S1 (a. a. residues 13 – 672) and S2 (a. a. residues 680 – 1192) were expressed (Broer *et al.*, 2006; Ho *et al.*, 2004a; Li *et al.*, 2003; Xiao *et al.*, 2003).

The expression levels and solubility of S1 and S2 may be low because of their high molecular weights (74 kDa and 56 kDa respectively). Another four domains with lower molecular weights were consequently selected for expression in parallel (Figure 2.1). One of these domains is the RBD (a. a. residues 318 – 510), which can be used to identify other interacting partner(s) other than ACE2, as well as act as a positive control. Another selected domain is a combination of HR1 and HR2. Finally, two fragments located at the upstream region of RBD and between RBD and HR1 were chosen.

2.1.2 Secondary Structure and Burial Region Predictions

Secondary structure and burial regions of S protein were predicted by an online database named "Jpred" (http://www.compbio.dundee.ac.uk/~www-jpred/) (Figure 2.2). The S protein a. a. sequence was also analysed by other algorithms including Network Protein Sequence Analysis (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html, http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html,http://npsa-pbil.ibcp.fr/cgibin/npsa_automat.pl?page=npsa_sopma.html), NNPREDICT(http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html), PORTER (http://distill.ucd.ie/porter/) to confirm the results. Three fragments (a. a. residues 15 – 317, 587 – 826 and 903 – 1187) were subsequently selected (Figure 2.1). Their N- and C-terminals contain random coils and are hydrophilic.

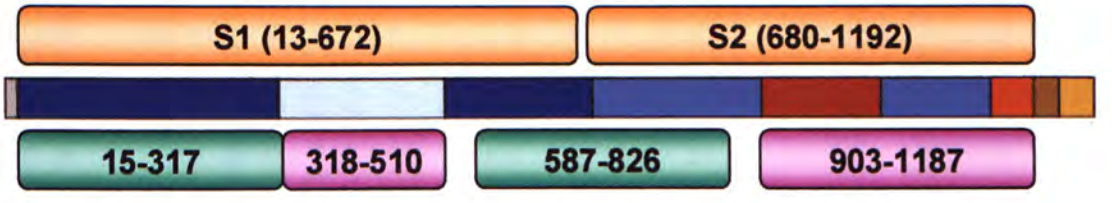
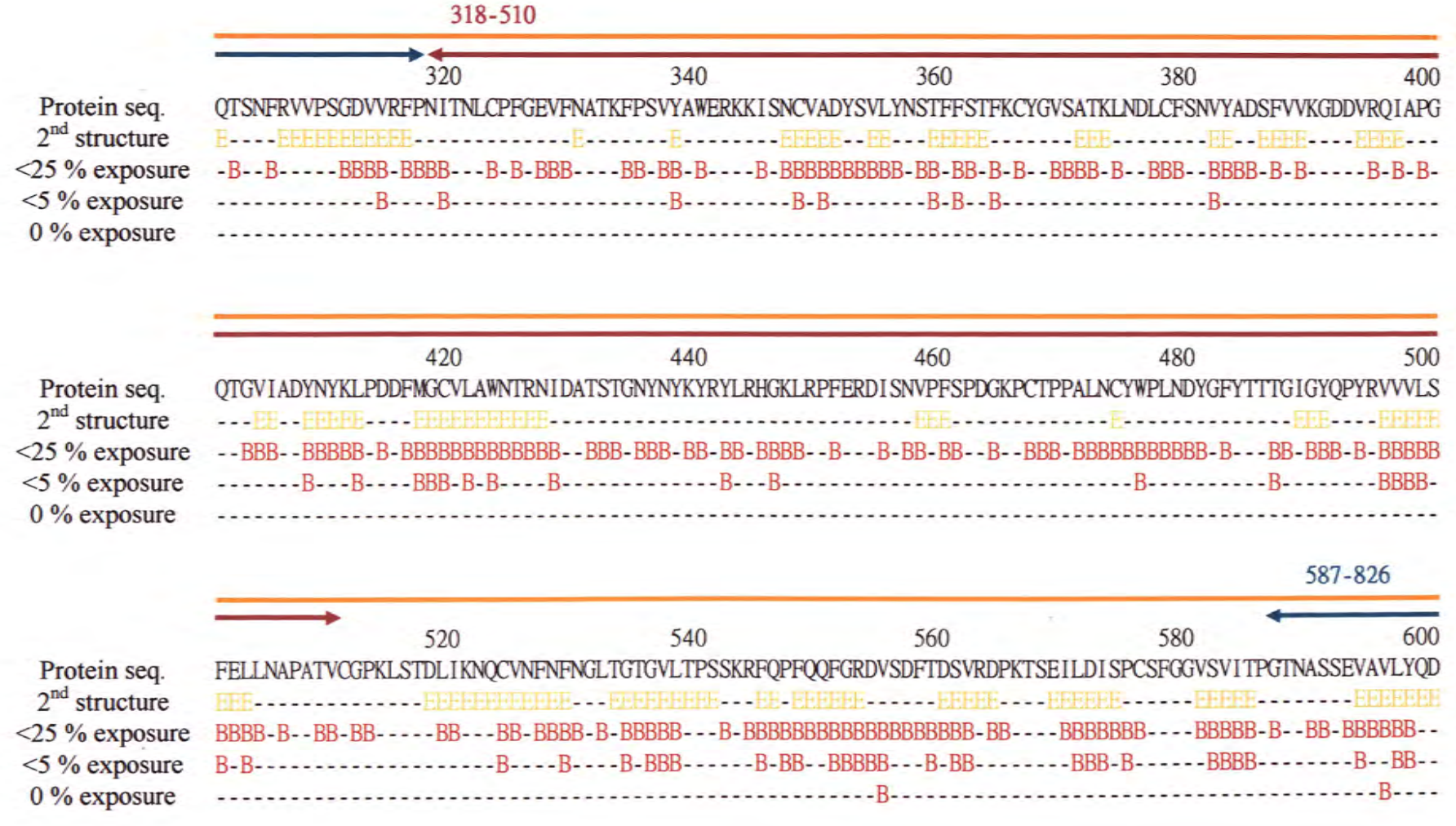
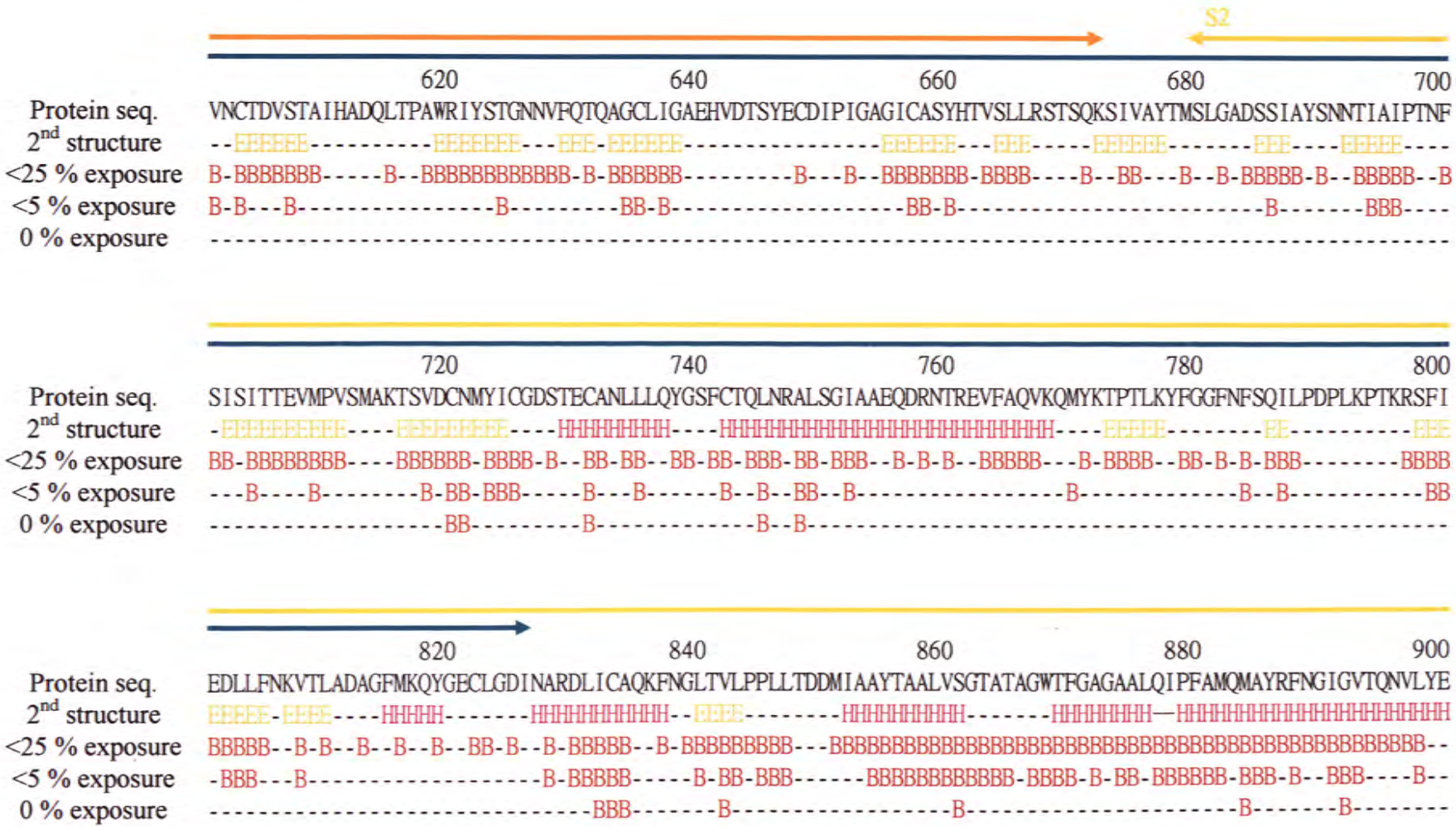


Figure 2.1 – Schematic diagram of selected fragments. Orange, green and purple rectangular blocks represent locations of S fragments.

Figure 2.2 – Secondary structure and burial region predictions of S glycoprotein. H and E in secondary structure prediction represent locations of α-helix and β-strand respectively, while B in burial prediction indicates chance of particular a.a. exposing to environment.


318-510
320
340
360
380
400
Protein seq.
QTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPG
2nd structure
<25 % exposure
<5 % exposure
0 % exposure
420
440
460
480
500
Protein seq.
QTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQPYRVVVLS
2nd structure
<25 % exposure
<5 % exposure
0 % exposure
587-826
520
540
560
580
600
Protein seq.
FELLNAPATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTNASSEVAVLYQD
2nd structure
<25 % exposure
<5 % exposure
0 % exposure


S2
620
640
660
680
700
Protein seq.
VNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYECDIPIGAGICASYHTVSLLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNF
2nd structure
<25 % exposure
<5 % exposure
0 % exposure
720
740
760
780
800
Protein seq.
SISITTEVMPVSMAKTSVDCNMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKTPTLKYFGGFNFSQILPDPLKPTKRSFI
2nd structure
<25 % exposure
<5 % exposure
0 % exposure
820
840
860
880
900
Protein seq.
EDLLFNKVTLADAGFMKQYGECLGDINARDLICAQKFNGLTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYE
2nd structure
<25 % exposure
<5 % exposure
0 % exposure


903-1187
920
940
960
980
1000
Protein seq.
NQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEI
2nd structure
<25 % exposure
<5 % exposure
0 % exposure
1020
1040
1060
1080
1100
Protein seq.
RASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVFNGTSWFITQRNFFSPQIITTD
2nd structure
<25 % exposure
<5 % exposure
0 % exposure
1120
1140
1160
1180
1200
Protein seq.
NTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWL
2nd structure
<25 % exposure
<5 % exposure
0 % exposure

```
                    1220                1240
Protein seq.        GFIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKFDEDDSEPVLKGVKLHYT
2nd structure       HHHHHHHHHHEEEEEEEEEE----------EEEEE---------
<25 % exposure      BBBBBBBBBBBBBBBBBBBBBBBBBB-BB--BB-BB--------BBB--B-BBBB
<5 % exposure       BBBB-BBBBBBBBBBBBBBBB------B----------------BB--B-----
0 % exposure        -B-------------------------------------------------------
```

2.2 Vector Amplification

2.2.1 E. coli Strain DH5α Competent Cell Preparation

Glycerol stock of *E. coli* strain DH5α was streaked on a Luria-Bertani (LB) plate (GE Healthcare) and incubated at 37 ℃ overnight. A single colony was inoculated to 5 mL of LB medium (GE Healthcare) and shaken at 37 ℃ overnight at 250 revolutions per minute (r.p.m.). The inoculum was transferred to 500 mL of LB medium, followed by shaking until an optical density at 600 nm (OD_{600}) between 0.4 – 0.6 was reached.

Harvested cell pellet was chilled on ice for 15 minutes in 167 mL of RF1 solution (pH 5.8) [100 mM rubidium chloride (RbCl), 50 mM manganese chloride, 30 mM potassium acetate, 10 mM calcium chloride ($CaCl_2$) and 25 mL of glycerol]. The harvesting and chilling procedures were repeated except RF1 solution was replaced by 40 mL of RF2 solution (pH 6.8) [10mM 3-(N-morpholino) propanesulfonic acid, 10 mM RbCl, 75 mM $CaCl_2$ and 6 mL of glycerol]. Aliquots of competent cells were stored at -80 ℃.

2.2.2 *Transformation of E. coli*

One microgram of pGEX-6P-1 (GE Healthcare), pET-28a (Novagen), pRHisMBP or pPICZα-A (Invitrogen) was added to 200 μL of thawed competent cells. After staying on ice for 15 minutes, cells were heat shocked at 42 ℃ for 90 seconds and then stayed on ice for five minutes. Next, 800 μL of LB medium was added and shaken at 37 ℃ for one hour at 250 r.p.m. for recovery. Then, 200 μL of them containing pGEX-6P-1 and pRHisMBP were spread to LB plates with 100 μg/mL of ampicillin (GE Healthcare) while that containing pET-28a and

pPICZα-A were spread to the plates with 30 μg/mL of kanamycin (Invitrogen) and 25 μg/mL of zeocin (Invitrogen) respectively. The plates were incubated at 37 ℃ overnight.

2.2.3 Small-scale Vector Amplification

Vector amplification was performed by "Rapid Miniprep System (250)" (Marlign biosciences). For each vector, a transformant on the plate was inoculated in 6 mL of LB medium with corresponding antibiotic and then shaken at 37 ℃ overnight at 250 r.p.m. Each harvested cell pellet was resuspended in 500 μL of Cell Suspension Buffer G1, followed by mixing with 500 μL of Cell Lysis Solution G2. After staying for five minutes, 350 μL of Neutralization Buffer M3 was mixed and centrifuged at 12,000 g for ten minutes. Supernatant was transferred to a column and centrifuged at 12,000 g for two minutes, followed by 800 μL of Wash Buffer G4. The vectors were eluted by 50 μL of distilled water (H_2O).

OD_{260} and OD_{280} of samples diluted in 100-folds were measured by using distilled H_2O as blank. The concentrations were calculated as follows: "Concentration (ng / μL) = OD_{260} x 100 (dilution fold) x 50" while the purities were calculated by OD_{260} / OD_{280} ratio.

2.3 Cloning of DNA Fragments into Various Vectors

2.3.1 Primer Design

All primers were designed by "GeneTool Lite" (http://www.DoubleTwist.com) using SARS-CoV strain CUHK-Su10 as template (Tsui *et al.*, 2003). Their lengths and melting temperatures are approximately 20 base pair (bp) and 54℃ respectively. Nucleotide sequence "TTA" ("TAA" in reverse and complement) encoding a stop codon for translation termination was added in each reverse primer involved in cloning into pGEX-6P-1 and pRHisMBP. Moreover, a cap sequence "TAGGGC" and a restriction site were provided at the beginning of each primer for restriction enzyme binding and digestion respectively (Table 2.1).

Table 2.1 – Features of S protein fragments.

a. a. residues of protein fragments	Range of nucleotides in the complete genome	Range of nucleotides in the S glycoprotein	Lengths of nucleotide sequence
13-672 (S1)	21513 – 23492	37 – 2016	1980
680-1192 (S2)	23516 – 25052	2038 – 3576	1539
15-317	21521 – 22427	43 – 951	909
318-510	22428 – 23006	952 – 1530	579
587-826	23235 – 23954	1759 – 2478	720
903-1187	24183 – 25037	2707 – 3561	855

Table 2.2 – Features of primers for cloning. Primers for cloning into pRHisMBP were identical to that for pGEX-6P-1. Italic, bold and underlined words represent cap, restriction site and stop codon respectively.

Name of plasmid	Primer sequence	Restriction enzyme	Melting temperature (℃)	Length (bp)
pGEX-S1	*TAGGGC***GGATCC**GGTAGTGACCTTGACCGGTG	*Bam*HI	64	20
	*TAGGGC***GAATTC**TTATTTTTGGCTAGTACTACGTAAT	*Eco*RI	58	22
pGEX-S2	*TAGGGC***GGATCC**TCTTTAGGTGCTGATAGTTCAAT	*Bam*HI	62	23
	*TAGGGC***GAATTC**TTATTA AATATATTGCTCATATTTTCCCA	*Eco*RI	58	23
pGEX-(15-317)	*TAGGGC***GGATCC**GACCTTGACCGGTGCA	*Bam*HI	52	16
	*TAGGGC***CTCGAG**TTAAGGGAATCTCACAACATCTC	*Xho*I	58	20
pGEX-(318-510)	*TAGGGC***GGATCC**AATATTACAAACTTGTGTCCTTTT	*Bam*HI	43	24
	*TAGGGC***CTCGAG**TTATTA AACCGTGGCCGGTG	*Xho*I	43	14
pGEX-(587-826)	*TAGGGC***GGATCC**GGAACAAATGCTTCATCTG	*Bam*HI	54	19
	*TAGGGC***CTCGAG**TTAAATATCACCTAGGCATTCG	*Xho*I	54	19
pGEX-(903-1187)	*TAGGGC***GGATCC**AAACAAATCGCCAACCA	*Bam*HI	48	17
	*TAGGGC***CTCGAG**TTATTTTCCCAATTCTTGAAGG	*Xho*I	52	19
pET-S1	*TAGGGC***GGATCC**GGTAGTGACCTTGACCGGTG	*Bam*HI	64	20
	*TAGGGC***CTCGAG**TTTTTGGCTAGTACTACGTAAT	*Xho*I	58	22
pET-S2	*TAGGGC***GGATCC**TCTTTAGGTGCTGATAGTTCAAT	*Bam*HI	62	23
	*TAGGGC***CTCGAG**TTA AATATATTGCTCATATTTTCCCA	*Xho*I	58	23

Name of plasmid	Primer sequence	Restriction enzyme	Melting temperature (℃)	Length (bp)
pET-(15-317)	*TAGGGC***GGATCC**GACCTTGACCGGTGCA	*Bam*HI	52	16
	*TAGGGC***CTCGAG**AGGGAATCTCACAACATCTC	*Xho*I	58	20
pET-(318-510)	*TAGGGC***GGATCC**AATATTACAAACTTGTGTCCTTTT	*Bam*HI	43	24
	*TAGGGC***CTCGAG**TTA AACCGTGGCCGGTG	*Xho*I	43	14
pET-(587-826)	*TAGGGC***GGATCC**GGAACAAATGCTTCATCTG	*Bam*HI	54	19
	*TAGGGC***CTCGAG**AATATCACCTAGGCATTCG	*Xho*I	54	19
pET-(903-1187)	*TAGGGC***GGATCC**AAACAAATCGCCAACCA	*Bam*HI	48	17
	*TAGGGC***CTCGAG**TTTTCCCAATTCTTGAAGG	*Xho*I	52	19
pPICZα-S1	*TAGGGC***GAATTC**GGTAGTGACCTTGACCGGTG	*Eco*RI	64	20
	*TAGGGC***GCGGCCGC**TTTTTGGCTAGTACTACGTAAT	*Not*I	58	22
pPICZα-S2	*TAGGGC***GAATTC**TCTTTAGGTGCTGATAGTTCAAT	*Eco*RI	62	23
	*TAGGGC***GCGGCCGC**TTAAATATATTGCTCATATTTTCCCA	*Not*I	58	23
pPICZα-(15-317)	*TAGGGC***GAATTC**GACCTTGACCGGTGCA	*Eco*RI	52	16
	*TAGGGC***GCGGCCGC**AGGGAATCTCACAACATCTC	*Not*I	58	20
pPICZα-(318-510)	*TAGGGC***GAATTC**AATATTACAAACTTGTGTCCTTTT	*Eco*RI	43	24
	*TAGGGC***GCGGCCGC**TTAAACCGTGGCCGGTG	*Not*I	43	14
pPICZα-(587-826)	*TAGGGC***GAATTC**GGAACAAATGCTTCATCTG	*Eco*RI	54	19
	*TAGGGC***GCGGCCGC**AATATCACCTAGGCATTCG	*Not*I	54	19
pPICZα-(903-1187)	*TAGGGC***GAATTC**AAACAAATCGCCAACCA	*Eco*RI	48	17
	*TAGGGC***GCGGCCGC**TTTTCCCAATTCTTGAAGG	*Not*I	52	19

2.3.2 DNA Amplification

DNA for pGEX-S1, pGEX-S2 and pGEX-(587-826) construction were amplified by polymerase chain reaction (PCR) using complementary deoxyribonucleic acid (cDNA) of CUHK-Su10 as template. 50 μL of each PCR mixture containing 0.2 mM deoxynucleoside triphosphate (dNTP), one unit of *pfu* polymerase and 1 X of *pfu* polymerase buffer (Promega), 2 μM of corresponding forward and reverse primers (Proligo), and 1 ng of cDNA were prepared. PCR mixture without template was prepared in parallel as a negative control. The cycling conditions for PCR were as follows: 94 ℃ for three minutes, 35 cycles of 94 ℃ for 36 seconds, 55 ℃ for 45 seconds, 72 ℃ for five minutes, and 72 ℃ for ten minutes as the final extension. The PCR was performed in DNA engine thermocycler (MJ Research) and amplified PCR products were used as templates to construct other plasmids.

2.3.3 DNA Purification

The PCR products were subjected to Tris(hydroxymethyl)aminomethane-acetate-ethylenediaminetetraacetic acid (TAE) agarose gel electrophoresis. 50 μL of each PCR mixture mixing with 10 μL of 6 X DNA loading dye [0.25 % (w/v) bromophenol blue, 0.25 % (w/v) xylene cyanol FF and 30 % (v/v) glycerol] was loaded to wells of the gel [40 mM Tris(hydroxymethyl)aminomethane-Hydrochloride (Tris-HCl), 20 mM acetic acid, 1 mM ethylenediaminetetraacetic acid (EDTA) (pH 8.0), 0.005 % (v/v) ethidium bromide and 1 % – 1.5 % (w/v) agarose)]. 2 μg of 100 bp ladder DNA marker or GeneRuler 1 Kb DNA ladder (Fermentas) was loaded in parallel to determine sizes of the PCR products. After electrophoresis,

the PCR products in the gel were visualized under UV light at an emission wavelength of 340 nm.

The regions of the gels containing the desired PCR products with expected sizes were isolated and extracted by "QIAQuick Gel Extraction Kit" (Qiagen). Each isolated gel slice was dissolved in 900 μL of buffer QC and then mixed with 300 μL of isopropanol. The mixture was transferred to a QIAquick Spin Column and centrifuged at 12,000 g for two minutes, followed by 750 μL of buffer PE. The PCR products were eluted by adding 30 μL of sterile H_2O at 70 ℃ and centrifuging at 12,000 g for three minutes.

2.3.4 Restriction Enzyme Digestion, Ligation and Transformation

Each mixture containing 28 μL of each purified PCR product, 2 X of One-phor-All buffer (GE Healthcare), and restriction enzymes (2 μL each), was prepared. Single-cut vector control and uncut vector control were prepared in parallel. The mixtures were digested at 37 ℃ for five hours and then heat inactivated at 85 ℃ for 15 minutes. Afterwards, purification of the restricted PCR products and vectors were performed as mentioned in section 2.3.2 (p.35).

Ten microliters of each mixture containing 6 – 7.5 μL of the digested PCR products, 1 – 2.5 μL of linearized vectors, 200 U of T4 DNA ligase and 1 X T4 DNA ligase buffer (New England BioLabs) were prepared. The ratio of the amount of PCR product and vector was approximately three to one. The mixture was ligated at 4 ℃ overnight and then transformed into DH5α competent cells.

2.3.5 Colony PCR

Six to thirty colonies were streaked to another LB plate with corresponding antibiotic and then transferred to 5 μL of PCR mixture containing 10 mM Tris-HCl (pH 9.0), 50 mM potassium chloride (KCl), 1.25 mM magnesium chloride ($MgCl_2$), 0.2 mM dNTP, 2 μM corresponding vector forward and reverse primers and 0.25 U of *Taq* polymerase. PCR was performed under the following cycling profile: 94 ℃ for three minutes, 35 cycles of 94 ℃ for 36 seconds, 55 ℃ for 45 seconds, 72 ℃ for two minutes, and 72 ℃ for ten minutes for the final extension. The presence of PCR products was determined by TAE agarose gel electrophoresis. For each construct, two transformants were selected for small-scale plasmid amplification.

2.4 DNA Sequence Analysis

2.4.1 Primer Design

Six vector primers were purchased according to instructions of manufacturers. Internal primers were designed according to the genome sequence of SARS-CoV strain CUHK-Su10. The distance between internal primers was approximately 600 bp (Table 2.3).

Table 2.3 – Features of primers for DNA sequence analysis.

Name of primer	Primer sequence	Location in S protein	Length (bp)
pGEX-F (pGEX)	GGCAAGCCACGTTTGGTG	/	18
T7-P (pET / pRHisMBP)	TAATACGACTCACTATA	/	17
α-F (pPICZα)	TACTATTGCCAGCATTGCTGC	/	21
37BF	AACCCATGGGTACACAGACA	415-434	20
38BF	GAAAAAAAATTTCTAATTGT	1015-1034	20
39BF	CAAAGAGATTTCAACCATTT	1615-1634	20
40BF	GCACACAACTAAATCGTGCA	2215-2234	20
41AF	TCTGCTCACTGATGATATGA	2525-2544	20
42AF	ACATGTCACGTATGTGCCAT	3125-3144	20
pGEX-R(pGEX)	GAGCTGCATGTGTCAGAGG	/	19
T7-T (pET / pRHisMBP)	GCTAGTTATTGCTCAGC	/	17
AOX1-R (pPICZα)	GCAAATGGCATTCTGACATCC	/	21
37AR	TTTTCTGAAACATCAAGCGA	395-414	20
38AR	TTCAACTTAGTGGCAGAAAC	1095-1114	20
39AR	GGTGAAATGTCTAATATTTC	1695-1714	20
39BR	TTTGGCTAGTACTACGTAAT	1985-2004	20
40BR	CACCAAATGTCCATCCAGCA	2585-2604	20
41BR	GGAAGTATGCTTTGCCTTCA	3185-3204	20

2.4.2 DNA Amplification and Purification for DNA Sequence Analysis

Twenty microliters of each PCR mixture containing 2 μL of BigDye Terminator v3.1 Cycle Sequencing RR-100 (Applied Biosystems), 0.75 X of BigDye Terminator v1.1, v3.1 Sequencing Buffer (Applied Biosystems), 0.5 μM vector or gene-specific primer, and 200 ng of plasmid was prepared. The primers for DNA sequence analysis could be found in table 2.4. The cycling conditions for PCR were as follows: 96 ℃ for three minutes, 26 cycles of 96 ℃ for ten seconds, 55 ℃ for five seconds and 60 ℃ for four minutes.

Each PCR product mixing with 5 μL of 125 mM EDTA and 60 μL of 99.9 % (v/v) ethanol (EtOH) was centrifuged at 1,500 g for 30 minutes at 4 ℃. Supernatant was removed by centrifugation at 50 g for one minute at 4 ℃. Each pellet resuspended in 60 μL of 70 % EtOH was further centrifuged at 1,500 g for 15 minutes at 4 ℃. After repeating the centrifugation, each pellet was resuspended in 15 μL of Hi-Di Formamide (Applied Biosystems) and denatured at 95 ℃ for two minutes.

Table 2.4 – Primers for recombinant plasmid sequence analysis.

Name of plasmid	Name of primer
pGEX-S1	pGEX-F, 37BF, 38BF, 39BF, pGEX-R, 37AR, 38AR, 39AR
pGEX-S2	pGEX-F, 41AF, 42AF, pGEX-R, 40BR, 41BR
pGEX-(15-317)	pGEX-F, 37BF, pGEX-R, 37AR
pGEX-(318-510)	pGEX-F, pGEX-R
pGEX-(587-826)	pGEX-F, 40BF, pGEX-R, 39BR
pGEX-(903-1187)	pGEX-F, 42AF, pGEX-R, 41BR
pET-S1	T7-P, 37BF, 38BF, 39BF, T7-T, 37AR, 38AR, 39AR
pET-S2	T7-P, 41AF, 42AF, T7-T, 40BR, 41BR
pET-(15-317)	T7-P, 37BF, T7-T, 37AR
pET-(318-510)	T7-P, T7-T
pET-(587-826)	T7-P, 40BF, T7-T, 39BR
pET-(903-1187)	T7-P, 42AF, T7-T, 41BR
pRHisMBP-S1	T7-P, 37BF, 38BF, 39BF, T7-T, 37AR, 38AR, 39AR
pRHisMBP-S2	T7-P, 41AF, 42AF, T7-T, 40BR, 41BR
pRHisMBP-(15-317)	T7-P, 37BF, T7-T, 37AR
pRHisMBP-(318-510)	T7-P, T7-T
pRHisMBP-(587-826)	T7-P, 40BF, T7-T, 39BR
pRHisMBP-(903-1187)	T7-P, 42AF, T7-T, 41BR
pPICZα-S1	α-F, 37BF, 38BF, 39BF, *AOX1*-R, 37AR, 38AR, 39AR
pPICZα-S2	α-F, 41AF, 42AF, *AOX1*-R, 40BR, 41BR
pPICZα-(15-317)	α-F, 37BF, *AOX1*-R, 37AR
pPICZα-(318-510)	α-F, *AOX1*-R
pPICZα-(587-826)	α-F, 40BF, *AOX1*-R, 39BR
pPICZα-(903-1187)	α-F, 42AF, *AOX1*-R, 41BR

2.4.3 Sequence Detection and Result Analysis

The sequencing products were subsequently analysed by the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). The results were recorded by "ABI PRISM 3100/3100-Avant Data Collection", generated in chromatogram format and visualized by "Chromas Version 2.13" (http://www.mb.mahidol.ac.th/pub/chromas/chromas.htm). The nucleotide sequences were then analysed by the "nucleotide-nucleotide BLAST" under the "Basic Local Alignment Search Tool" (BLAST) developed by the Nation Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/) for alignment.

Protein Expression, Purification and Analysis

3.1 Protein Expression in *E. coli*

3.1.1 Molecular Weight and pI Predictions

The molecular weights (M. W.) and pI of S protein fragments with various tags were predicted by "Compute pI / Mw tool" (http://au.expasy.org/tools/pi_tool.html) in order to determine the integrity of expressed proteins and the influence of buffer pH on protein solubility.

3.1.2 Glycerol Stock Preparation

One transformant containing each constructed plasmid was inoculated in 0.5 mL of LB medium with corresponding antibiotic. Each inoculum was mixed with 0.5 mL of sterile glycerol and stored at -80 ℃.

3.1.3 Protein Expression Induction

One microliter of Rosetta 2, C41, Origami (kindly provided by Professor W. N. Au.) and BL21 glycerol stock carrying plasmid was inoculated in LB medium with corresponding antibiotic and then shaken at 37 ℃ overnight at 250 r.p.m. The inoculum was transferred to LB medium with corresponding antibiotic so that the final overnight culture percentage was 1 % and further shaken at 37 ℃ at 250 r.p.m. until mid-log phase (OD_{600} = 0.4 – 0.6) was reached. Iso-propylthio-β-D-galactopyranoside (IPTG) (Shanghai Shenergy Biocolor Bioscience & Technology Company) was added so that its final concentration was 0.2 mM. The cultures were shaken at 250 r.p.m. for 3 – 16 hours.

3.1.4 Protein Extraction

Cells expressing Glutathione S-Transferase (GST) fusion protein were harvested and cell pellet was resuspended in GST buffer [10 mM Tris-HCl (pH 7.4), 300 mM sodium chloride (NaCl), 0.5 mM EDTA, 2 mM dithiothreitol (DTT) and 0.2 mM phenylmethylsulfonyl fluoride (PMSF)] or phosphate buffer saline (PBS) with 2 mM DTT and 0.2 mM PMSF. Addictives including 100 mM DTT, 1 M urea and 1 % (v/v) triton X-100 (Sigma) were individually added to GST buffer during optimization. On the other hand, cell pellet expressing His fusion proteins or His-Mannose binding protein (MBP) fusion protein was resuspended in His buffer [10 mM Tris-HCl (pH 7.4), 250 mM NaCl, and 0.2 mM PMSF] with 5 mM imidazole (Sigma).

The resuspended cells were disrupted by using the Sonoplus ultrasonic homogenizer system (Bandeln). The sonication processes were performed on ice to prevent over-heating. 10 μL of lysate was saved as total lysate for analysis. The lysate left was centrifuged at 48,000 g for 30 minutes at 4 ℃ and supernatant containing soluble fusion protein was stored at 4 ℃. 10 μL of supernatant was saved for subsequent analysis.

3.1.5 Affinity Chromatography

Five milliliters of supernatant containing soluble GST fusion protein was added to 100 μL of bed volume of equilibrated glutathione sepharose (GE Healthcare). The mixture in 1.5 mL of microcentrifuge tube was gently shaken at 4 ℃ for one hour so that the GST fusion protein was immobilized on the glutathione sepharose. After centrifugation, 10 μL of flow through was saved for analysis while the remaining was discarded. The glutathione sepharose was washed

five times with 1 mL each of buffer to remove non-specifically bound proteins. The immobilized GST fusion protein was finally eluted three times with 100 μL each of buffer containing 10 mM reduced glutathione (Calbiochem). Eluents were mixed and stored at 4 ℃.

For His fusion proteins as well as His-MBP fusion proteins, the procedures were same as that for GST fusion protein except His-binding resin (Merck) were used instead of glutathione sepharose. After charging of His-binding resin by 50 mM nickel sulphate ($NiSO_4$), His buffers with 5 mM, 50 mM and 500 mM imidazole were used to equilibrate the resin, wash and elute the proteins respectively.

3.1.6 Removal of GroEL

To remove GroEL by adenosine triphosphate (ATP), 150 mL of BL21 expressing GST-S1 at 16 ℃ overnight was lyzed in 5 mL of GST buffer. After extraction, supernatant was added to 5 mL of 2 X ATP buffer (pH 7.4) (GST buffer with 20 mM ATP disodium salt (Sigma), 40 mM $MgCl_2$ and 100 mM KCl). The mixture was gently shaken at 4 ℃, 25 ℃ or 37 ℃ overnight. Then, 100 μL of bed volume of equilibrated glutathione sepharose was added and gently shaken for one hour. After immobilization, the glutathione sepharose was washed five times with 1 mL each of 1 X ATP buffer (pH 7.4). GST-S1 was finally eluted three times with 50 μL of 1 X ATP buffer with 10 mM reduced glutathione. Eluents were mixed and stored at 4 ℃.

On the other hand, to remove GroEL by urea, 200 mL of BL21 expressing GST-S1 at 16 ℃ overnight was lyzed in 12 mL of PBS. After extraction, supernatant with soluble GST-S1 was added to 450 μL of bed volume of equilibrated glutathione

sepharose. The mixture was gently shaken at 4 ℃ for one hour and then washed five times by 1 mL each of PBS. 50 μL of the glutathione sepharose was gently shaken in 1 mL of PBS with 0 M, 0.5 M, 1 M, 1.25 M, 1.5 M, 1.75 M, 2 M, 2.25 M or 2.5 M urea at 4 ℃ overnight. After washing three times by 1 mL each of PBS with various urea concentrations followed by PBS without urea, the immobilized GST-S1 samples were finally eluted three times with 100 μL each of PBS with 10 mM reduced glutathione. Eluents were mixed and stored at 4 ℃.

3.1.7 Protein Solubilization and Refolding

His-(15-317), His-(318-510), His-(587-826), His-(903-1187) were individually expressed in 50 mL of BL21 cells at 25 ℃ for four hours and then extracted. Afterwards, each insoluble cell pellet was resuspended in 5 mL of solubilization buffer [His buffer with 6 M guanidine hydrochloride (guHCl)] with 5 mM imidazole, followed by centrifugation at 48,000 g for 30 minutes at 4 ℃. Each His-fusion protein in the supernatant was immobilized on 500 μL of His-binding resin, washed five times by 1 mL each of solubilization buffer with 50 mM imidazole and eluted three times by 150 μL of solubilization buffer with 500 mM imidazole.

The unfolded recombinant proteins were refolded in buffers with various conditions by rapid dilution. In brief, 2 μL of each unfolded His-fusion protein in a concentration of about 0.5 μg/μL was added to 190 μL of refolding buffers and mixed immediately. The procedures were repeated four times so that the final volume was 200 μL. After staying at 4 ℃ overnight, the mixture was centrifuged at 14,000 g for 15 minutes at 4 ℃ in order to remove insoluble protein and the supernatant was then analysed by SDS-PAGE. After primary screening, the refolding

condition with the highest refolding yield was chosen for further secondary and tertiary screenings. The compositions of refolding buffers were listed in table 3.1.

Table 3.1 – Compositions of refolding buffers. All buffers for primary and secondary screenings contain 1 mM EDTA. L-arginine, reduced glutathione (GSH), oxidized glutathione (GSSG), polyethylene glycol (PEG), $MgCl_2$, $CaCl_2$ and EDTA were addictives of buffers.

	No.	gu-HCl (M)	L-arginine (mM)	GSH (mM)	GSSG (mM)
Primary screening	1	0	0	2	0.2
	2	0	0.4	2	0.4
	3	0	0.8	1	1
	4	0.5	0	2	0.4
	5	0.5	0.4	1	1
	6	0.5	0.8	2	0.2
	7	1	0	1	1
	8	1	0.4	2	0.2
	9	1	0.8	2	0.4
	No.		PEG (μM)		
Secondary screening	1	The best refolding condition in primary screening	0		
	2		50		
	3		100		
	4		150		
	5		200		
	6		250		
	No.		$MgCl_2$ / $CaCl_2$ (mM)	EDTA (mM)	
Tertiary screening	1	The best refolding condition in secondary screening	0 / 0	1	
	2		2 / 2	0	
	3		4 / 4	0	
	4		6 / 6	0	
	5		8 / 8	0	
	6		10 / 10	0	

3.2 Protein Expression in *P. pastoris*

3.2.1 Large-scale Plasmid Amplification

Plasmid amplification was performed by "HiSpeed Plamid Midi Kit" (Qiagen). 1 μL of DH5α glycerol stock containing the plasmid was inoculated in 200 mL of LB medium with 25 μg/mL of zeocin and shaken at 37 ℃ overnight at 250 r.p.m. Harvested bacterial pellet was resuspended in 7 mL of Buffer P1, followed by 7 mL of Buffer P2. After incubating for five minutes, 7 mL of Buffer P3 was added and the mixture was poured immediately into the barrel of the QIAfilter Cartridge followed by ten minutes incubation. In the meanwhile, HiSpeed Midi Tip was equilibrated in 4 mL of Buffer QBT. Afterwards, the cap from the QIAfilter Cartridge was removed and a plunger was gently inserted so that the cleared cell lysate passed through the resin in the equilibrated HiSpeed Tip. The HiSpeed Tip was washed by 20 mL of Buffer QC and the plasmid was eluted by 5 mL of Buffer QF.

The eluent mixing with 3.5 mL of isopropanol was incubated for five minutes for DNA precipitation. The mixture was transferred into a 20 mL syringe, in which the outlet nozzle was attached to the QIAprecipitator Midi Module, and then filtered through the QIAprecipitator using constant pressure so that the plasmid was trapped. Afterwards, the QIAprecipitator was removed from the 20 mL syringe and the plunger was pulled out. After re-attaching the QIAprecipitator, the plasmid was washed by 2 mL of 70 % EtOH by inserting and pressing the plunger using constant pressure. Finally, the QIAprecipitator was attached onto a new 5 mL syringe and then the plasmid was eluted by 1 mL of distilled H_2O. Plasmid concentration and purity were measured.

3.2.2 Restriction Enzyme Digestion and Ethanol Precipitation

A restriction mixture containing 25 μg of plasmid, 6 μL of *Sac*I and 1 X of the buffer L (GE Healthcare) was incubated at 37 ℃ for five hours, followed by heat inactivation at 85 ℃ for 15 minutes. The mixture was mixed with 99.9 % EtOH until the final percentage of EtOH reached 70 %. After standing in -20 ℃ for 15 minutes, the plasmid was harvested by centrifuging at 12,000 g for 15 minutes at 4 ℃. Supernatant was discarded and pellet was dried completely by vacuum and then resuspended in 12 μL of sterile H_2O. Plasmid concentrations were determined by measuring OD_{260} and adjusted to 1 μg/μL. 12 μL of sample containing 0.5 μg of plasmid and 1 X DNA loading dye was analysed by TAE agarose gel electrophoresis to confirm the linearization of plasmid.

3.2.3 Preparation of KM71H Competent Cells

P. pastoris strain KM71H glycerol stock was streaked on a Yeast Extract Peptone Dextrose (YPD) plate [1 % (w/v) yeast extract (GE Healthcare), 2 % peptone (w/v) (Becton, Dickinson and Company), 2 % (w/v) glucose and 2 % (w/v) agar] and then incubated at 28 ℃ until colonies formed. A single colony inoculated in 10 mL of YPD medium was shaken at 28 ℃ overnight at 250 r.p.m. until OD_{600} at 2 – 6 was reached. 1.5 mL of the culture was harvested and resuspended in 1.5 mL of sterile H_2O, followed by 0.75 mL of H_2O, 0.5 mL and 50 μL of 1 M sorbitol (Sigma) at 4 ℃ respectively.

3.2.4 Electroporation

Eighty microliter of KM71H competent cells mixing with 10 μL of linearized plasmid were transferred to Eppendorf Electroporation Cuvette with 2 mm gap width (Eppendorf) and then incubated on ice for five minutes. The cuvette was pulsed by Eppendorf electroporator 2510 (Eppendorf) with the following parameters: 5 kV/cm, 10 μF, 600 Ω and 5 ms. 1 mL of 1 M sorbitol was immediately added and the mixture was incubated at 28 ℃ for 90 minutes. It was then spreaded on Yeast Extract Peptone Dextrose Sorbitol (YPDS) plate (YPD plate with 1 M sorbitol and 100 μg/mL of zeocin). The YPDS plates were incubated at 28 ℃ for three to ten days until colonies formed.

3.2.5 Colony PCR

The procedures mentioned in section 2.3.5 (p.37) was performed except KM71H transformants were streaked to another YPDS plate and incubated at 28 ℃ for three to ten days until colonies formed.

3.2.6 Protein Expression Induction and Time Course Study

A KM71H trasnformant was inoculated in 1 mL of Buffered Glycerol-Complex Medium (BMGY) (pH 6.0) [1 % yeast extract, 2 % peptone, 1.34 % (w/v) yeast nitrogen base without amino acid (Sigma), 132mM dibasic potassium phosphate, 868mM monobasic potassium phosphate, 0.00004 % (w/v) biotin and 1 % glycerol] and then shaken at 28 ℃ overnight at 250 r.p.m. The inoculum was sub-cultured to 100 mL of BMGY and further shaken at 28 ℃ at 250 r.p.m. overnight until mid-log phase (OD_{600} = 2 – 6) was reached.

Harvested yeast cells were resuspended in 10 mL of Buffered Methanol-Complex Medium (BMMY) (pH 6.0) containing BMGY except glycerol was replaced by 0.5 % (v/v) methanol (MeOH). The cultures were shaken at 28 ℃ for ten days at 250 r.p.m. for protein expression induction. 1 mL of culture was saved everyday and stored at -20 ℃ for expression level analysis. Final percentage of 0.5 % MeOH was added everyday to compensate evaporation. Recombinant protein in medium was purified by affinity chromatography.

3.2.7 Deglycosylation

A mixture containing the recombinant protein, 1 X of G7-buffer, 1 % (v/v) NP-40 and 0.2 μL of Peptide-N-glycanase F (PNGase F) (New England BioLabs) was digested at 37 ℃ for one hour. Deglycosylated protein was analysed by SDS-PAGE.

3.3 Protein Analysis

3.3.1 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

Ten microliters of total lysate, supernatant and flow through and 20 μL of eluent in 1 X SDS Loading dye [50 mM Tris (pH 6.8), 100 mM DTT, 2 % (w/v) sodium dodecyl sulfate (SDS), 0.1 % bromophenol blue and 5 % glycerol] were boiled at 100 ℃ for five minutes for protein denaturation. In the meanwhile, 10 %, 12 % or 15 % SDS-polyacrylamide gel of 1 mm thick was prepared.

The samples were loaded to wells of the SDS-polyacrylamide gel. At the same time, 2 μL of high range protein markers (GE Healthcare) or low range prointe markers (GE Healthcare) in 1 X SDS Loading dye were loaded in parallel. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed at a voltage of 180 V until dye front was just departed from the gel. The gel was stained in staining solution [45 % EtOH, 10 % (v/v) acetic acid and 0.24 % bromophenol blue], followed by destaining in detaining solution (30 % EtOH and 10 % acetic acid) until background colour was removed.

In order to determine the target protein amount, various amounts of bovine serum albumin (BSA) were subjected to SDS-PAGE. The target protein amount was estimated by comparing band intensities.

3.3.2 Western Blotting

After SDS-PAGE, proteins in the gel without staining were transferred to the PVDF transfer membrane, which was equilibrated by 70 % EtOH followed by transfer buffer (25 mM Tris-HCl, 200 mM glycine and 10 % MeOH). Transfer was performed using the Semi-dry electroblotter (OWL separation systems) at a voltage of 15 V for 75 minutes.

The membrane was immersed in PBS with 0.1 % (v/v) polyoxyethylenesorbitan monolaurate (PBST) with 5 % (w/v) skimmed milk powder to block unbounded area. Then, it was probed by immersing in 5 mL of PBS-T with 5 % milk powder and primary antibody with gently shaking at 4 ℃ overnight. The antibody was either mouse monoclonal anti-GST antibody (Santa Cruz Biotechnology) or anti-His antibody (Santa Curz Biotechnology) in dilutions of 1:1000 and 1:200 respectively. After washing five times with PBST, the membrane was probed by gently shaking in 5 mL of PBS-T with 5 % milk and anti-mouse horseradish peroxidase (HRP)-linked secondary antibody in a dilution of 1:2500 at room temperature for one hour. After further washing for five times, the membrane was immersed in 150 μL of Enhanced Luminol Reagent and 150 μL of Oxidizing Reagent of Western Lightning Chemiluminescence Reagent (PerkinElmer). Signals were captured by Fuji Super Rx Film (FUJI Photo products Co. Ltd.).

3.3.3 Mass Spectrometry

After SDS-PAGE, selected target bands were isolated and sliced into small pieces. Each sample was gently shaken in 1 mL of coomassie blue destaining solution [75 mM ammonium bicarbonate (NH_4HCO_3) and 40 % EtOH]. After destaining and washing by H_2O, the sample was incubated two times in 100 μL of 200 mM NH_4HCO_3 for ten minutes, followed by 100 μL of acetanitrate (ACN) for five minutes. After completely dried by vacuum, the gel was stayed in 10 μL of trypsin solution [200 ng of trypsin (Promega) and 50 mM $(NH_4)_2CO_3$] on ice for 30 minutes. 50 mM $(NH_4)_2CO_3$ was added until the gel was just covered and incubated at 30 ℃ overnight for trypsin digestion. Next, it was sonicated for ten

minutes in the presence of 20 μL of 50 mM $(NH_4)_2CO_3$. 2 μL of sample was transferred to the Matrix-Assisted Laser Desorption Ionisation (MALDI) Sample Plate. Detailed procedure of the mass spectrometric analysis can be found in sections 4.4.4 and 4.4.5 (p.59).

3.3.4 N-terminal Sequencing

One hundred picomoles (3 μg) of recombinant protein was transferred to a PVDF membrane. The membrane was stained by coomassie blue staining solution [0.5 % (w/v) coomassie brilliant blue, 40 % MeOH and 10 % acetic acid] for 30 seconds and then destained by distilled H_2O. The protein on the membrane was isolated and sequenced by Precise Peptide Sequencing System 492.

3.3.5 Size Exclusion Chromatography

The eluent was further purified by using ATKA Prime Plus (GE Healthcare) connecting to equilibrated HiLoad 16/60 Superdex 200 preparative grade column (GE Healthcare). After filtering and injection, the sample was passed through the column with a flow rate of 3 mL/min. 2 mL of each fraction was collected while real-time OD_{280} was recorded by "PrimeView 5.0" (GE Healthcare). The size of the target protein was determined by comparing the elution volumes of the target protein and standard proteins including Blue dextran (Void volume), Ferritin (440 kDa), Ovalbumin (43 kDa) and Ribonuclease A (13.7 kDa) (GE Healthcare). Finally, the proteins in different fractions were analysed by SDS-PAGE.

Identification of Interacting Partner(s)

4.1 Vero E6 Cell Preparation

4.1.1 Cell Culture

Vero E6 cells originated from fetal rhesus monkey kidney cells are SARS-CoV susceptible. They were cultivated in 150 mm flasks in Dulbecco's Modified Eagle Medium (DMEM) (pH 7.4) with 10 % (v/v) fetal bovine serum (Invitrogen Life Technology) and 5 % (v/v) penicillin-streptomycin (Invitrogen Life Technology) at 37 ℃ with the supply of 5 % (v/v) carbon dioxide. To prepare Vero E6 total lysate, cells of 70 % – 90 % confluence were washed by PBS and detached by trypsin digestion (Invitrogen Life Technology). Approximately 90 % Vero E6 cells were harvested by centrifugation at 1,000 g for three minutes. After washing with PBS, the harvested cells were stored at -80 ℃. The remaining cells were used further for propagation.

4.1.2 Protein Extraction and Western Blotting

To screen interacting partners, 1 g of Vero E6 cells with approximately 5 x 10^9 cells were lysed by 10 mL of mammalian cell lysis buffer [50 mM Tris (pH 7.4), 70 mM NaCl, 1 % triton X-100 and one-fifth of protease inhibitor cocktail tablet (Roche)]. After shaking gently for two hours at 4 ℃, the lysate was centrifuged at 1,000 g for five minutes at 4 ℃ and then the supernatant was stored at 4 ℃ for pull-down assay. 1 μL of total lysate, supernatant and pellet were analysed by SDS-PAGE. The membrane was probed by mouse monoclonal anti-hACE2 antibody (R&D systems) at the dilution of 1:200.

4.2 Pull-down Assay

One hundred micrograms of purified GST-S1 and 50 μg of purified GST mixed with 600 μL and 300 μL of equilibrated glutathione sepharose respectively were gently shaken at 4 ℃ for one hour. To screen Vero E6 interacting partners, GST-S1 and GST immobilized on glutathione sepharose were washed by 1 mL and 0.5 mL of mammalian cell lysis buffer respectively. Afterwards, the glutathione sepharose bound to GST-S1 was divided into two portions. 5 mL of lysis buffer with soluble Vero E6 lysate was added to one portion while 5 mL of lysis buffer without proteins was added to another portion. Another 5 mL of lysis buffer with soluble lysate was added to glutathione sepharose bound with GST. The samples with bound GST and without lysate acted as negative controls to identify non-specific binding proteins (Table 4.1). After gently shaking at 4 ℃ for two hours, they were washed for five times with 1 mL each of mammalian cell lysis buffer. GST-fusion proteins with interacting partners were eluted three times by 100 μL of mammalian cell lysis buffer with 10 mM reduced glutathione. Eluents were pooled together for future analysis.

Table 4.1 – Putative proteins presented in eluents of samples and negative controls. ✓ represents proteins that are only present in eluent of GST-S1 with Vero E6 proteins. x represents proteins are present in the eluents of negative controls as well.

Protein presented in eluent	GST-S + Vero E6	GST-S	GST + Vero E6
GST-S	x	x	
GST tag			x
Bacterial proteins binding to S	x	x	
Bacterial proteins binding to GST	x	x	x
Bacterial proteins binding to resin	x	x	x
Vero E6 proteins binding to S	✓		
Vero E6 proteins binding to GST	x		x
Vero E6 proteins binding to resin	x		x
Vero E6 proteins binding to bacterial proteins that bound to S	✓		
Vero E6 proteins binding to bacterial proteins that bound to GST	x		x
Vero E6 proteins binding to bacterial proteins that bound to resin	x		x

4.3 Two-dimensional Gel Electrophoresis

4.3.1 Isoelectric Focusing

Proteins in the eluent of mammalian cell lysis buffer were precipitated by providing nine volume of 99.9 % EtOH followed by centrifuging at 14,000 g for 15 minutes at 4 ℃. Each sample in form of pellet was resuspended in 250 μL of rehydration solution [8 M urea, 0.5 % (w/v) CHAPS (GE Healthcare), 0.2 % (w/v) DTT, 0.5 % (v/v) Immobilized pH gradient (IPG) buffer (GE Healthcare) and 0.002 % bromophenol blue]

After centrifugation at 12,000 g for ten minutes, each supernatant was applied into a 13 cm strip holder (GE Healthcare). A 13 cm pH 3 – 10 strip (GE Healthcare) was placed with the gel side down. The anodic end of the strip was positioned toward the pointed end of the strip holder. After removing bubbles, the strip was entirely covered by IPG cover fluid (GE Healthcare). Isoelectric focusing (IEF), the first dimensional gel electrophoresis, was performed by Ettan IPGphor III isoelectric unit (GE Healthcare) with the following: 30 V for ten hours or more, 500 V for one hour, 1,000 V for one hour, and then at 8,000 V (or maximal voltage) until 16,000 Vhr was reached. The maximum current of electrophoresis per strip was 50 μA. The strip after electrophoresis was stored in -80 ℃ or immediately prepared for the second dimensional gel electrophoresis.

4.3.2 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

Before SDS-PAGE, each strip was equilibrated in 10 mL of equilibration solution (2 % SDS, 50 mM Tris (pH 8.8), 6 M Urea, 30 % glycerol and 0.002 % bromophenol blue) with 100 mg of DTT for 15 minutes with gently shaking, followed by 10 mL of equilibration solution with 250 mg of Iodoacetamide (IAA).

Meanwhile, a 12 % SDS-polyacrylamide gel, which was 14 cm width, 16 cm height and 1.5 mm thick, was prepared for each strip. Gel solution was applied until approximately 5 mm height of space was left.

The equilibrated strip was placed on the top of the SDS-polyacrylamide gel so that the strip of the gel did not touch the glass. A 5 mm wide filter paper containing 2 μL of high range molecular weight marker or full range marker (Invitrogen) was placed in parallel. The strip and the paper were covered by adding SDS-running buffer (25 mM Tris-HCl, 250 mM glycine, 0.1 % SDS with 0.5 % agarose and 0.002 % bromophenol blue). SDS-PAGE was performed at a voltage of 180 V until the tracking dye was just departed from the gel.

4.3.3 Silver Staining

The gels were stained by silver staining kit (GE Healthcare). Each gel was shaken in 250 mL of fixation solution (40 % EtOH and 10 % acetic acid) for 30 minutes followed by incubation in 250 mL of sensitizing solution [30 % EtOH, 0.2 % (w/v) sodium thiosulphate, and 17 g of sodium acetate]. After washing three times by 250 mL each of H_2O for five minutes, the gel was shaken in 250 mL of silver reaction solution [0.25 % (w/v) silver nitrate solution and 0.148 % (w/v) formaldehyde] for 20 minutes. After washing twice by 250 mL each of H_2O for one minute, the gel was shaken in 250 mL of developing solution (6.25 g of sodium carbonate and 0.0074 % formaldehyde) until dark spots were developed. The developing reaction was stopped by immersing the gel in 250 mL of stopping solution (3.25 g of EDTA-Na_2-$2H_2O$) for ten minutes. Finally, the gel was stored in H_2O. Nano-pure H_2O was used in all procedures.

4.4 Mass Spectrometry

4.4.1 Destaining

Selected spots stained by silver staining were isolated and destained by 20 μL of solution (15 mM potassium ferricyanide and 50 mM sodium thiosulphate). When the color was entirely removed, the sample was washed by H_2O and then incubated twice in 100 μL of 200 mM $(NH_4)_2CO_3$ (Sigma) for ten minutes.

4.4.2 In-gel Digestion

The destained gel was dehydrated twice by 100 μL of ACN. After completely dried by vacuum, the gel was incubated in 10 μL of solution [200 ng of trypsin and 50 mM $(NH_4)_2CO_3$] on ice for 30 minutes. 50 mM $(NH_4)_2CO_3$ was added until the gel was just covered and the mixture was incubated at 30 ℃ overnight for trypsin digestion.

The digested gel was sonicated for ten minutes in the presence of 20 μL of 50 mM $(NH_4)_2CO_3$ and the supernatant was saved. The sonication procedures were repeated for three times except 50 mM $(NH_4)_2CO_3$ was replaced by a solution [50 % (v/v) ACN and 2.5 % (v/v) trifluoroacetic acid (TFA)]. Then, the sonication procedures were further repeated by using 10 μL of ACN. Finally, all stored supernatants were mixed and dried by vacuum.

4.4.3 Desalting by Zip-Tip

Each vacuum-dried sample was resuspended in 10 μL of 0.1 % TFA. Afterwards, by pipetting up and down slowly, Ziptip (Millipore) was wet twice by 10 μL of 50 % ACN and then equilibrated twice by 10 μL of 0.1 % TFA. Peptides were bound to the resin in the Ziptip by pipetting ten more times. After washing five times by 10 μL of 0.1 % TFA, the peptides in Ziptip were eluted by 2 μL of elution solution (50 % ACN and 0.1 % TFA).

4.4.4 Loading Sample

Elution solution containing peptides was transferred to 192 Well Stainless Steel MALDI Sample Plate (Applied Biosystems) for mass spectrometric analysis. Finally, 0.5 μL of matrix solution [10 mg/mL of cinnamic acid (Sigma), 50 % ACN and 0.1 % TFA] was added to dried pellets for crystallization.

4.4.5 Peptide Mass Detection and Data Analysis

Peptide masses were detected by 4700 Proteomics Discovery System (Applied Biosystems) including Matrix-Assisted Laser Desorption Ionisation-time of flight-time of flight (MALDI-TOF-TOF) mass spectrometer. The peptide mass fingerprints (PMFs) were submitted to online databases including NCBI, Swiss-Prot and TrEMBL for analysis via "GPS Explorer" as well as another two searching algorithms, including Mascot (http://www.matrixscience.com/search_form_select.html) and Aldente (http://www.expasy.org/tools/aldente). The search parameters were listed in table 4.2. If a known protein was analysed by mass spectrometry, its a. a. sequence was submitted to "Peptide mass" (http://ca.expasy.org/tools/peptide-mass.html) in order to predict PMFs.

Table 4.2 – Search parameter of PMFs.

Databases	NCBI / Swiss-Prot / TrEMBL
Species	All species, mammalian, bacteria, viruses
Peptide mass tolerance	100 ppm
Peptide masses	Monoisotopic, $[M+H]^+$
Enzyme	Trypsin
Allow for	One missed cleavage site
Mass exclusion list	Trypsin : 842.508 1045.564 2225.12 2299.176 897.531 1940.953 2239.105 2465.199 904.468 2211.105 2283.181 Keratin: 1011.500 1338.500 1699.825 2288.954 1037.527 1357.696 1707.773 2312.148 1060.564 1357.719 1716.852 2383.900 1066.499 1365.640 1838.915 2383.952 1179.601 1373.655 1993.977 2510.132 1184.591 1383.691 2083.017 2705.162 1193.617 1434.771 2093.087 2807.250 1234.680 1475.785 2221.100 2831.195 1296.685 1570.677 2230.105 3312.309 1307.678 1638.860 2230.170 3337.600 1320.583 1657.793 2239.175 3338.700 GST-S1: 927.5258 1140.5167 1310.6752 1779.8296 963.5298 1154.5629 1480.6914 1804.8864 1046.5629 1257.7313 1576.8733 2011.9946 1085.5374 1258.6102 1756.9915 2534.2272 GST: 1011.6309 1516.8039 2045.0338 2326.1403 1032.5870 1801.9476 2269.1386 2357.2063 1182.6841 GroEL: 875.4482 1201.6535 1580.3970 1845.9188 921.5040 1238.6600 1589.8897 2010.9510 985.5676 1291.6562 1605.8594 2399.2776 1000.5673 1373.7028 1640.8714 2632.3498 1011.5217 1385.7933 1668.8084 2851.3301 1045.5524 1454.6394 1711.7769 3239.7238 1154.5333 1567.8802 1783.0146 3322.4990 1199.6704

S Protein Expression

5.1 Plasmid Construction

Six PCR products encoding S protein fragments were amplified by PCR and examined by TAE agarose gel electrophoresis in order to clone into pPICZα-A (Figure 5.1). The results of the cloning of other PCR products into pGEX-6P-1, pET-28a and pRHisMBP were similar.

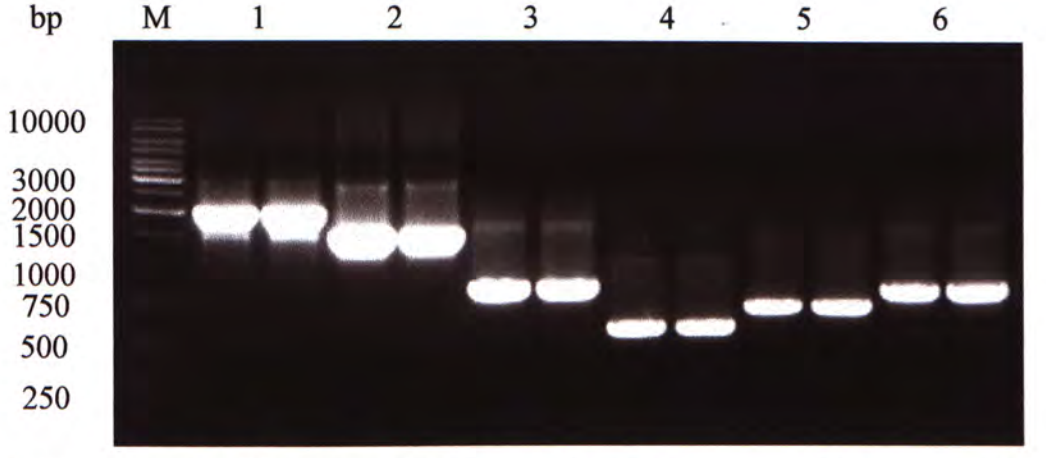


Figure 5.1 – Amplification of DNA encoding S glycoprotein fragments. Six PCR products were amplified and their sizes were same as expected. M: 1000bp marker. 1 – 6: PCR products encoding S1, S2, 15-317, 318-510, 587-826 and 903-1187 respectively.

After ligation and transformation, the presence of pPICZα-A-S1 in transformants was examined by colony PCR using vector primers, α-F and *AOX1*-R (Figure 5.2). The results of the cloning of other constructs were similar.

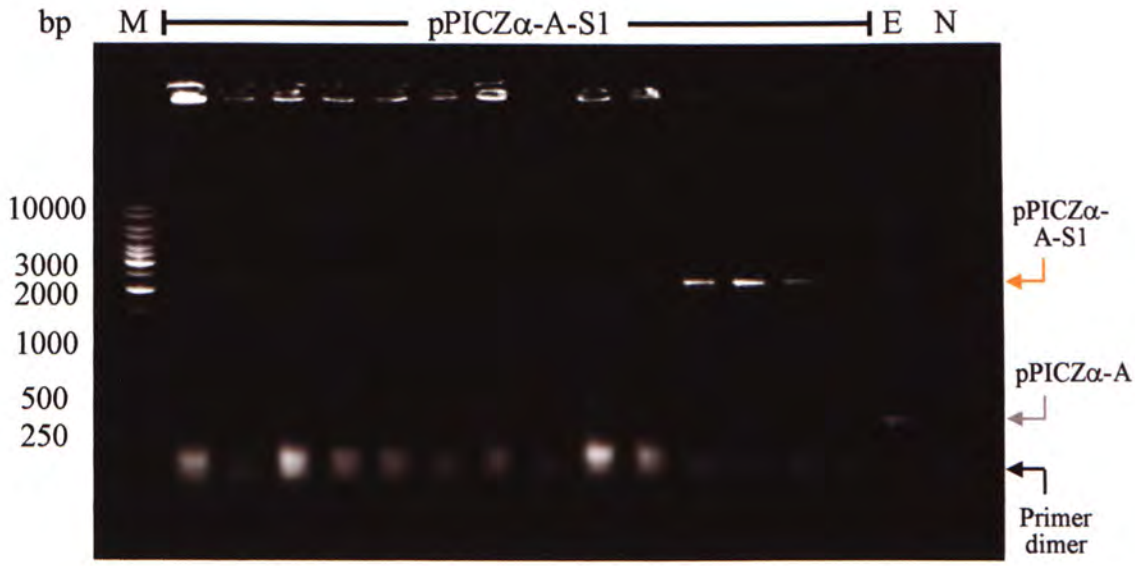


Figure 5.2 – Colony PCR of pPICZα-A-S1. 2.2 kb PCR products were specifically amplified by using a few of pPICZα-A-S1 transformants as template. M: 1000 bp marker. pPICZα-A-S1: Transformants as template. E: pPICZα-A as template. N: No template. The orange, grey and black arrows indicate expected locations of the PCR products of pPICZα-A-S1, pPICZα-A and primer-dimer.

Plasmids in at least two transformants were amplified by small–scale plasmid amplification. Their average concentrations and OD_{260}/OD_{280} ratios were approximately 100 ng/μL and 1.8 respectively. After sequence analysis, the plasmids with completely matched sequences were stored for protein expression.

5.2 Molecular Weight and pI Predictions

pI of most fusion proteins expressed in *E. coli* and *P. pastoris* were different from pH of buffers used (pH 7.4 and 6.0 respectively) (Table 5.1). Except αA-S1 and αA-(318-510), their differences were higher than 0.5. Thus the solubility of fusion proteins was not depleted due to pH environment.

Table 5.1 – Predicted molecular weights and pI of S protein fragments. The numbers without bracket were predicted M. W. in kDa while that with brackets were predicted pI. Bolded numbers represent the differences between pI and buffer pH was less than 0.5.

a. a. residues	Without tag	*E. coli*			*P. Pastoris*
		GST-tag	His-tag	His-MBP-tag	His-myc-tag
13-672 (S1)	73.540 (5.79)	100.364 (5.76)	78.149 (6.42)	116.350 (5.56)	**77.223 (5.55)**
680-1192 (S2)	56.399 (5.24)	83.223 (5.44)	61.008 (6.13)	99.296 (5.28)	60.093 (5.18)
15-317	41.909 (5.64)	61.083 (5.74)	38.868 (6.64)	77.068 (5.47)	37.967 (5.42)
318-510	21.840 (8.14)	48.664 (6.21)	26.449 (8.41)	64.649 (5.71)	**25.556 (5.78)**
587-826	26.068 (4.87)	52.983 (5.32)	30.677 (6.21)	68.878 (5.16)	29.781 (4.91)
903-1187	31.573 (5.59)	58.397 (5.66)	36.182 (6.51)	74.382 (5.42)	35.283 (5.36)

5.3 Protein Expression and Optimization in *E. coli*

5.3.1 Comparison of Expression Levels, Solubility and Purities of S Protein Fragments

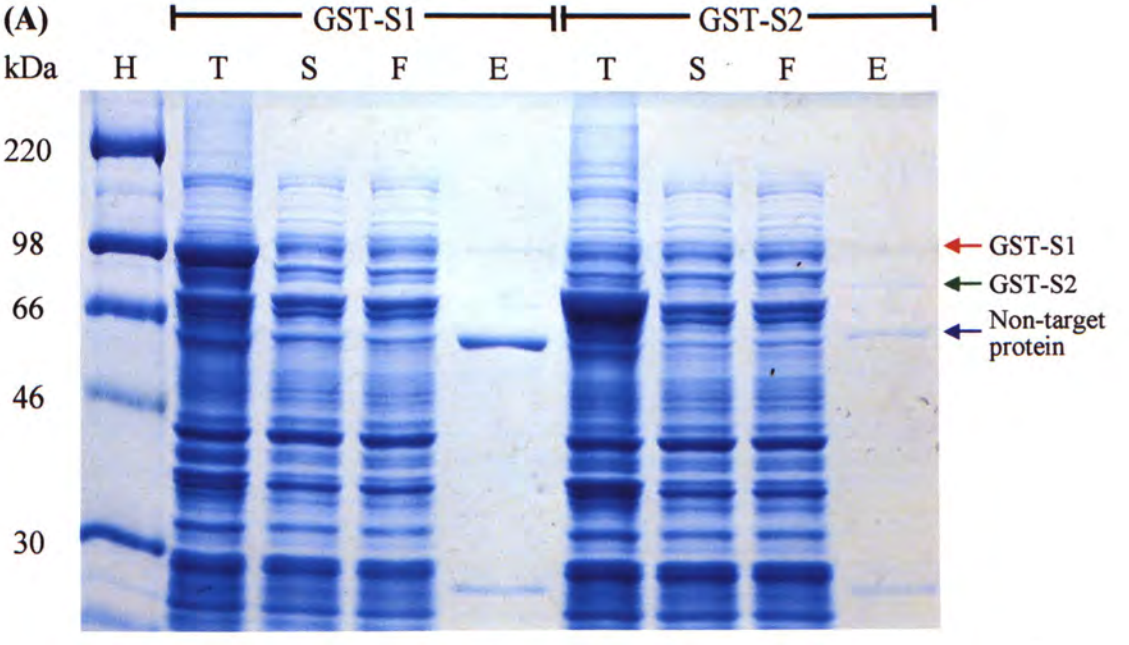
Expressions of six S fragments with GST-, His- and His-MBP-tags were induced by 0.2 mM of IPTG at mid-log phase and expressed in *E. coli* strain BL21 at 25 ℃ for four hours. After purification by affinity chromatography, total lysate, supernatant, flow through and eluent of each fragment were examined by SDS-PAGE in order to determine the expression level, solubility, binding affinity towards resin and purity respectively. The identities of proteins were confirmed by western blotting analysis.

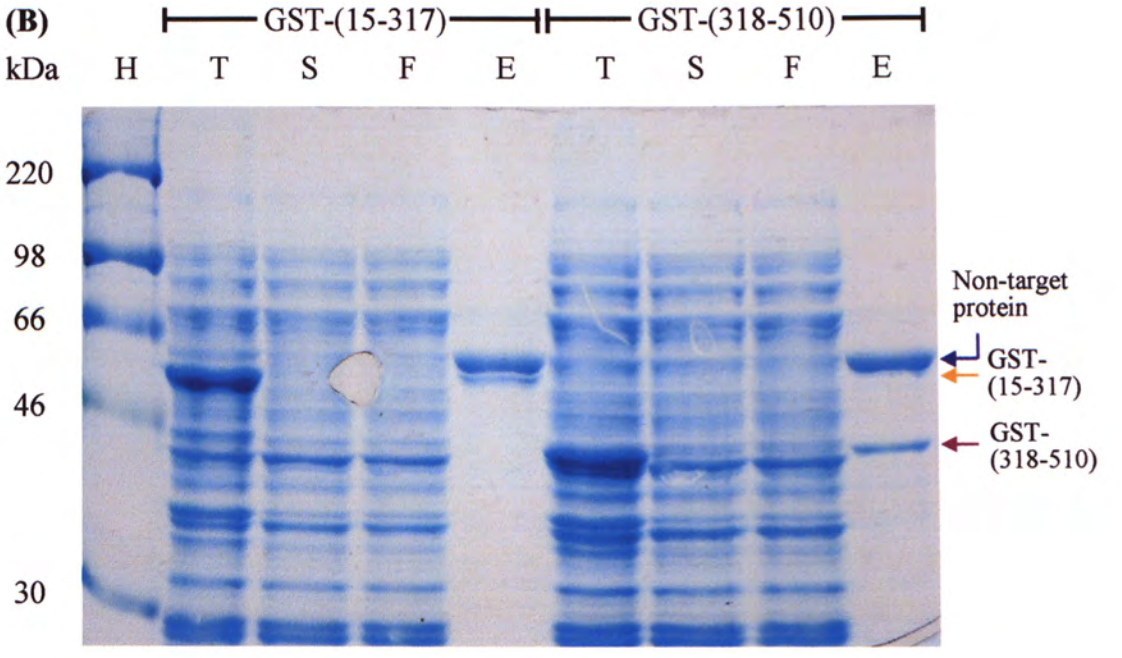
The expression levels and solubility of four short GST-fusion proteins were higher than that of GST-S2 while that of GST-S1 was the lowest (Figure 5.3 A-C). The yields of GST-fusion proteins purified from 50 mL of *E. coli* were 5μg or less (Table 5.2). However, the solubility and purity of proteins were generally low. Besides non-specific binding proteins, a protein of 60 kDa in size could not be removed by affinity chromatography.

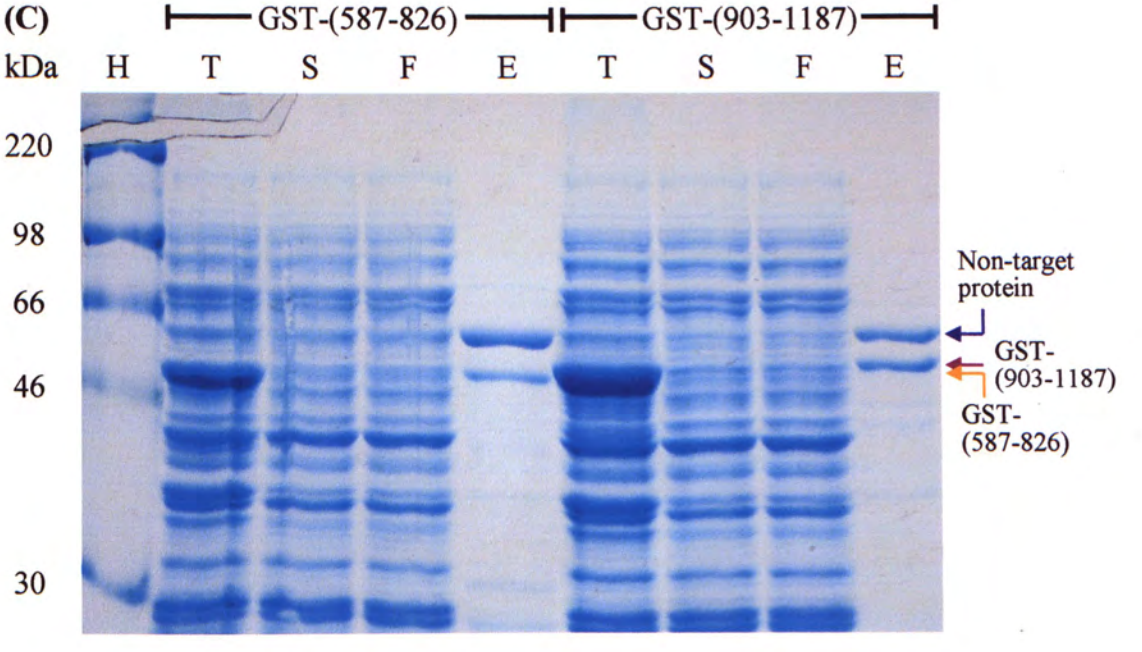
The six His-fusion proteins with expected sizes could be expressed but were completely insoluble (Figure 5.3 D and Table 5.2). On the other hand, the expression levels and solubility of His-MBP-fusion proteins were similar or higher than that of GST-fusion proteins (Table 5.2). Their purities were however much lower than that of GST-fusion proteins (Figure 5.3 A and E). Since S1 was predicted to interact with the receptor, GST-S1 was selected for the optimization of expression conditions and further purification.

Table 5.2 – The estimated amounts of partial purified S protein fragments. Protein expression was induced by 0.2 mM of IPTG at mid-log phase of *E. coli* strain BL21 and shaken at 25 ℃ for four hours. The protein amounts were estimated by comparing the band intensities with BSA of known amounts.

a. a. residues	GST-tag	His-tag	His-MBP-tag
13-672 (S1)	<10 μg/L	0 μg/L	<10 μg/L
680-1192 (S2)	<10 μg/L	0 μg/L	0 μg/L
15-317	100 μg/L	0 μg/L	250 μg/L
318-510	100 μg/L	0 μg/L	300 μg/L
587-826	100 μg/L	0 μg/L	600 μg/L
903-1187	100 μg/L	0 μg/L	1500 μg/L




(B)
GST-(15-317)
GST-(318-510)
kDa
H
T
S
F
E
T
S
F
E
220
98
66
46
30
Non-target protein
GST-(15-317)
GST-(318-510)


(C)
GST-(587-826)
GST-(903-1187)
kDa
H
T
S
F
E
T
S
F
E
220
98
66
46
30
Non-target protein
GST-(903-1187)
GST-(587-826)

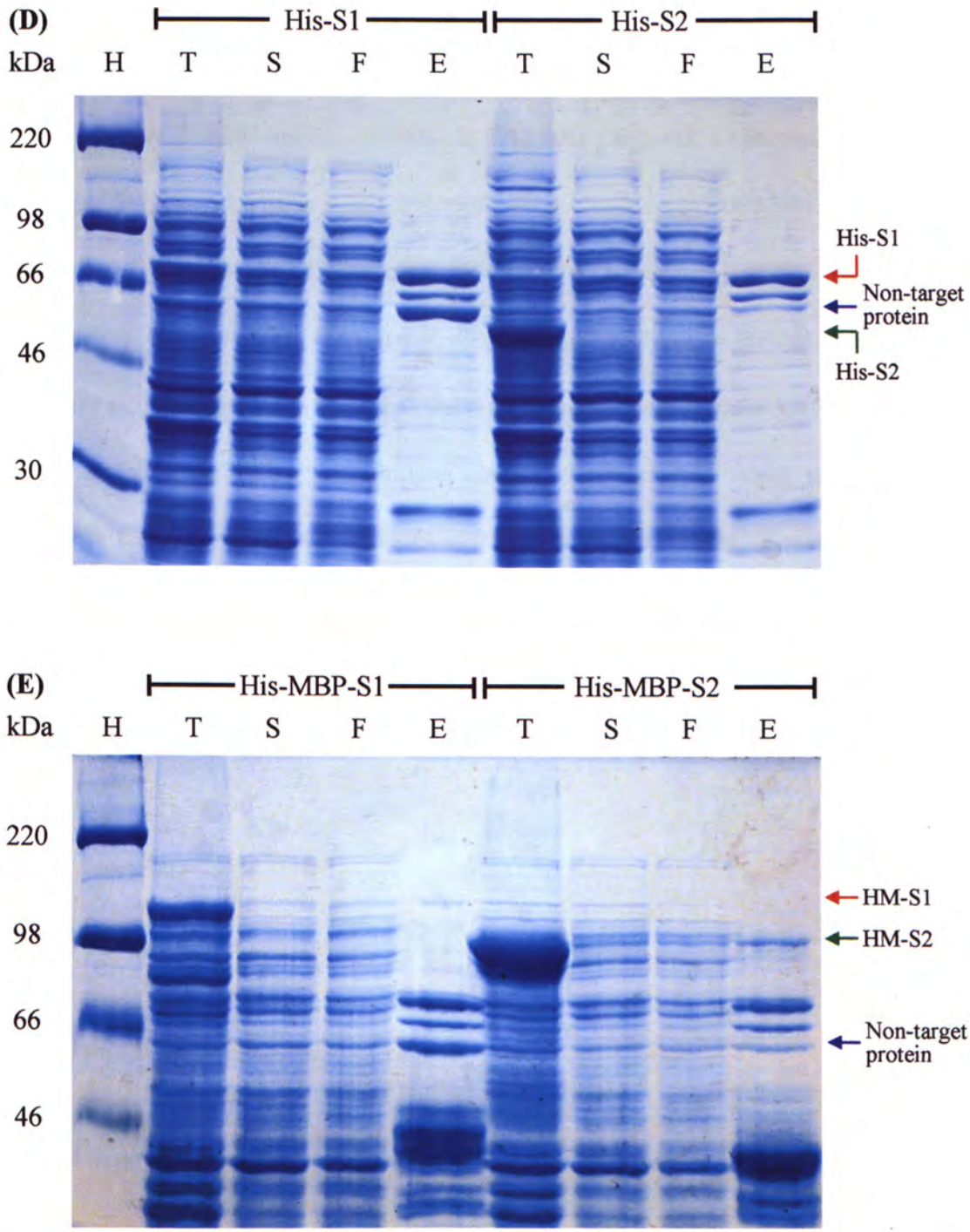


Figure 5.3 – Purification of S protein fragments with GST-, His- and His-MBP-tags. Target bands of **(A)** GST-S1, GST-S2, **(B)** GST-(13-317), GST-(318-510), **(C)** GST-(587-826) and GST-(903-1187) with expected sizes were observed in total lysates and eluents. A protein band of 60 kDa was detected in each eluent. **(D)** Target bands of His-S1 and His-S2 could be observed in total lysates but not eluents. **(E)** Results of S1 and S2 with His-MBP-tag were similar to that with GST-tag but a high proportion of unrelated bands were detected in eluents. H: High range marker. T: Total lysate. S: Supernatant. F: Flow through. E: Eluent. Red, green, orange, purple and blue arrows represent expected locations of S1, S2, GST-(15-317) and GST-(587-826), GST-(318-510) and GST-(903-1187), and the non-target protein respectively.

5.3.2 Alteration of the Soluble Protein Amount in Various Cell Strains, Expression Conditions and Lysis Buffers

Other *E. coli* strains, Rosetta 2, C41 and Origami, expressing GST-S1 in different expression temperatures and expression levels were examined. Nonetheless, amounts of soluble GST-S1 in other strains were similar, while that expressed at 37 ℃ were slightly lower than that at 25 ℃, suggesting that S protein expressing in a higher temperature leads to a high degree of degradation (Figure 5.4). On the other hand, no increase in soluble protein amount was detected when the expression duration was altered to three hours, six hours and overnight (Figure 5.4). Thus, the expression duration was not an important factor to control expression levels and soluble protein amounts. In conclusion, lowering expression temperature might be a possible method to raise the amounts of soluble GST-S proteins.

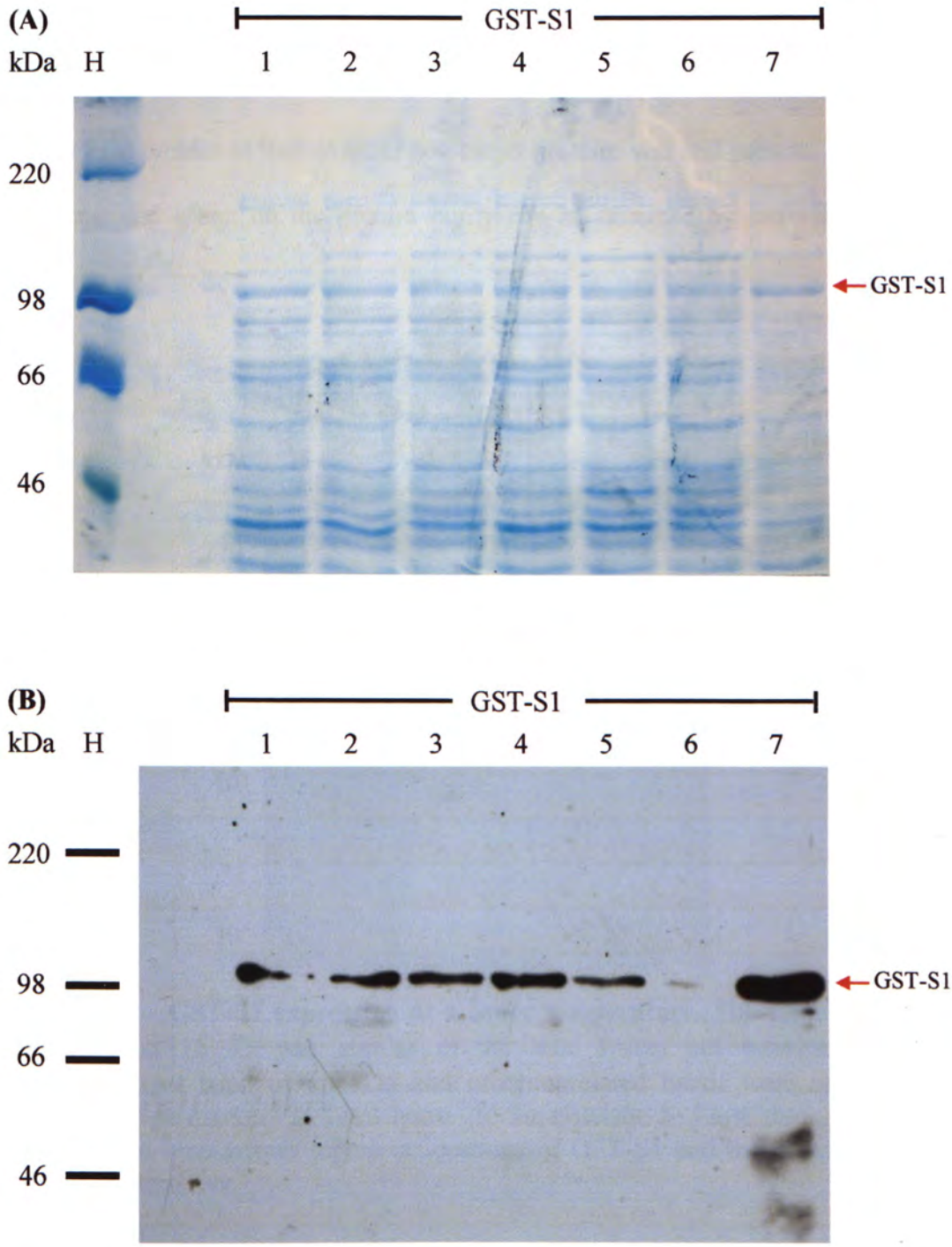


Figure 5.4 – GST-S1 expression at temperature and time. (A) SDS-PAGE and **(B)** western blotting showed that the soluble GST-S1 amounts expressed in various lengths of time were similar while that expressed at 37 ℃ were lower. The total lysate was presented in each western blotting analysis as internal control. H: High range marker. 1 – 3: Supernatant expressed at 25 ℃ for three hours, six hours and overnight. 4 – 6: supernatant expressed at 37 ℃ for three hours, six hours and overnight. 7: Total lysate at 25 ℃ for three hours. The red arrow represents GST-S1 location.

Solubility of GST-S1 expressed at 16 ℃ overnight was improved and the yield of purified GST-S1 was increased to 25 μg/L (Figure 5.5). However, the 60 kDa protein as well as other non-target proteins was still present, revealing that no beneficial effect on the protein purity can be obtained by lowering expression temperature.

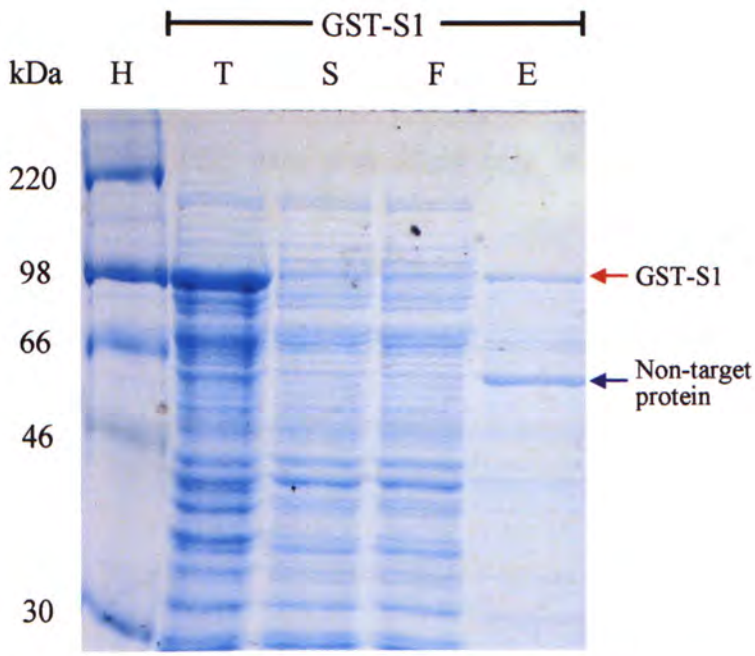


Figure 5.5 – GST-S1 expression at a lower temperature. The amount of GST-S1 expressed at 16 ℃ was similar in the total lysate but increased in eluent. The non-target band of 60 kDa and other unrelated bands were still presented. H: High range marker. T: Total lysate. S: Supernatant. F: Flow through. E: Eluent. The red and blue arrows represent locations of GST-S1 and the 60 kDa non-target protein respectively.

In order to further improve the solubility as well as the purity, *E. coli* expressing GST-S1 was lyzed by another four kinds of buffer: 1) GST buffer with 100 mM DTT; 2) GST buffer with 1 M Urea; 3) GST buffer with 1 % of Triton X-100; and 4) PBS. SDS-PAGE analysis showed that GST-S1 solubility was increased in three of these buffers (Figure 5.6). The amount of soluble GST-S1 lyzed in PBS was maximal, which was approximately five fold higher than that lyzed in GST-buffer. Furthermore, the 60 kDa non-target protein was reduced without reducing the amount of GST-S1 when *E. coli* was lysed in 1 M urea (Figure 5.6). These results indicated that PBS with urea might help in increasing the GST-S1 solubility and purity.

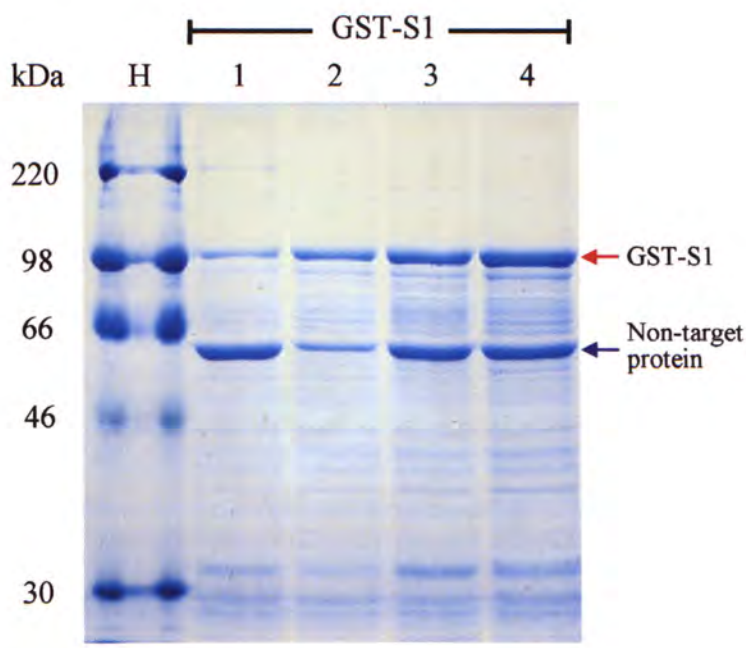


Figure 5.6 – Solubility of GST-S1 in various lysis buffers. The amounts of purified GST-S1 in the four buffers were shown. H: High range marker. 1: GST buffer. 2: GST buffer with 1 M urea. 3: GST buffer with 1 % Triton X-100. 4: PBS. The red and blue arrows represent locations of GST-S1 and the 60 kDa non-target protein respectively.

5.3.3 Identification and Removal of non-target proteins

Since the difference between molecular weights of GST-S1 and the non-target protein was 40 kDa, separation by size exclusion chromatography (SEC) was performed. GST-S1, which had been purified by affinity chromatography, was further purified by HiLoad 16/60 Superdex 200 column. They were nevertheless eluted together in the void volume (Figure 5.7). SDS-PAGE analysis showed both GST-S1 and GroEL were present in this peak, even their elution volumes in this peak were slightly varied, suggesting that they were bound together (Figure 5.8).

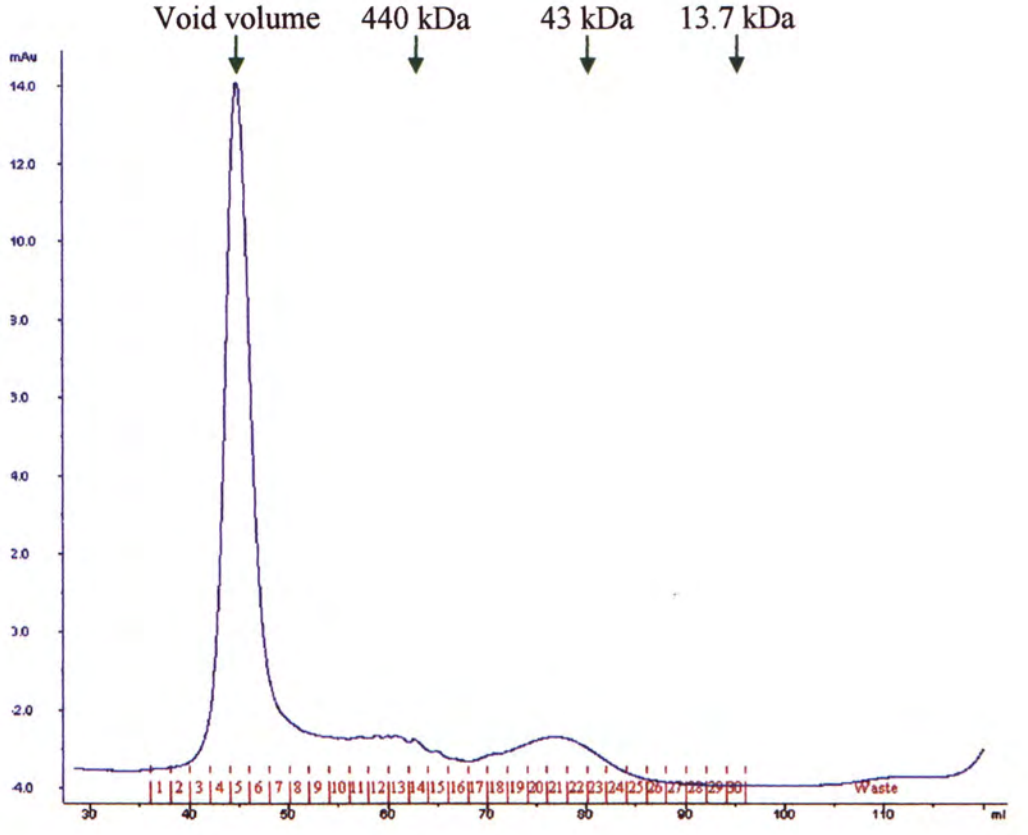


Figure 5.7 – Purification of GST-S1 by SEC. Only a single peak was eluted at 44 mL, which was the void volume by comparing with standards. Volume of each fraction was 2 mL. Green arrows indicate elution volumes of standards.

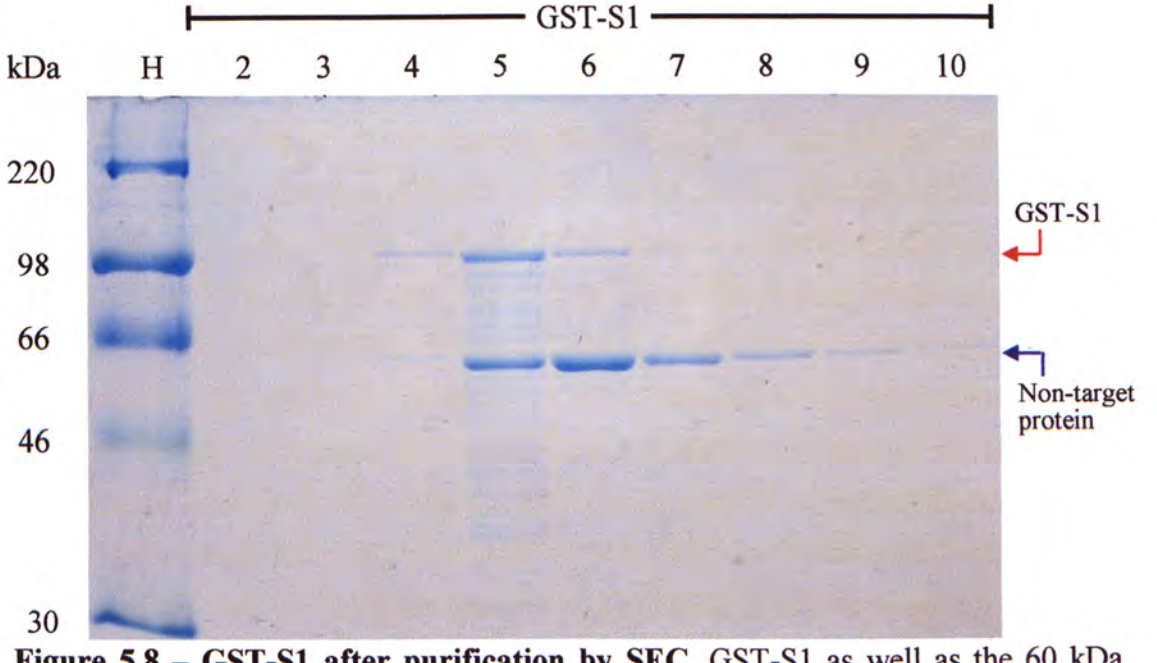


Figure 5.8 – GST-S1 after purification by SEC. GST-S1 as well as the 60 kDa non-target protein and little amount of other non-target proteins were eluted in the void volume. H: High range marker. Number: Fraction number of SEC on figure 5.7. The red and blue arrows represent locations of GST-S1 and the 60 kDa non-target protein respectively.

PMF showed the 60 kDa non-target band is a bacterial chaperone called GroEL (Figure 5.9 and Table 5.3). The role of this protein in bacterial is to help in folding of unfolded or misfolding proteins. It forms a 14-mers cage and binds to hydrophobic region of the unfolded or misfolded proteins (Landry *et al.*, 1994; Schmidt *et al.*, 1994; Van der Vaart *et al.*, 2004). By using ATP as substrate, the unfolded or misfolded protein is refolded and then released from the cage. Some proteins over-expressed in *E. coli* were bound by GroEL but released after ATP addition (Rohman *et al.*, 2000; Thomas *et al.*, 1997). However, in some cases, GroEL keeping protein in the native conformation was irremovable (Venkatesh *et al.*, 2004). As it was still unknown whether the binding of S to GroEL is beneficial, we tried to remove the GroEL by different methods. Since only a portion of GroEL can be removed in the presence of 1 M urea, completely removal of GroEL was tested by providing ATP and urea in various concentrations.

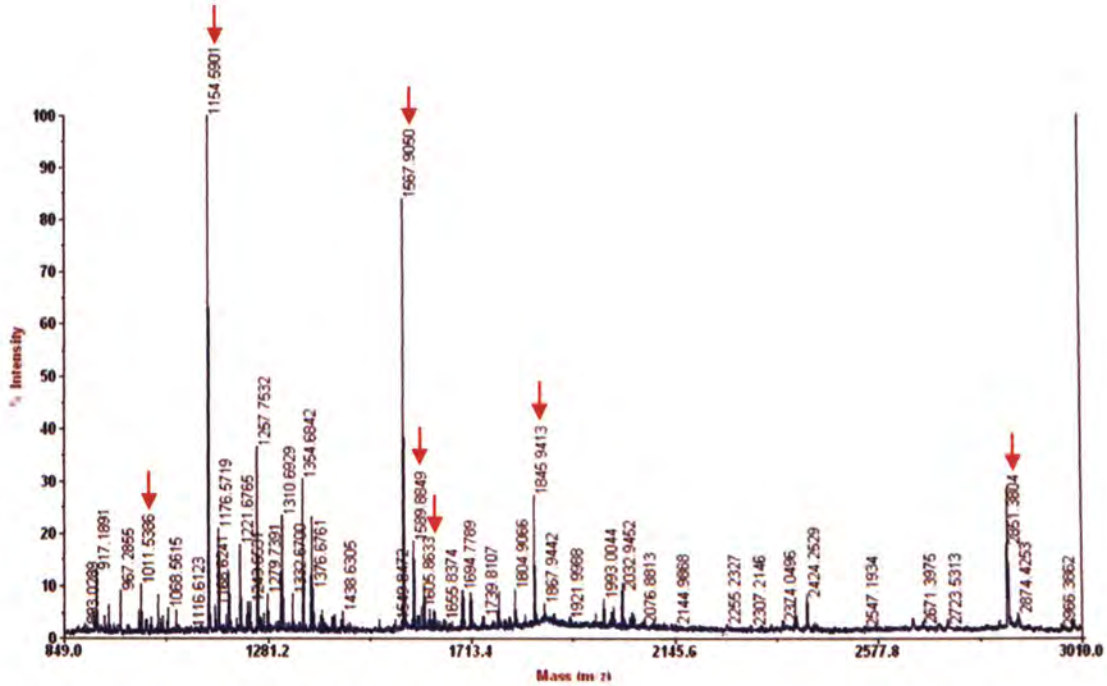


Figure 5.9 – PMF of the 60 kDa protein. The red arrows indicate peaks matching GroEL.

Table 5.3 – Peptides of the 60kDa protein matching to GroEL. Peptide masses were submitted to Mascot Search (http://www.matrixscience.com/search_form_select.html) for identification.

Start - End	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Sequence
396 - 404	1011.5386	1010.5313	1010.5145	0.0168	0	R.VEDALHATR.A
66 - 75	1154.5901	1153.5828	1153.5259	0.0569	0	K.FEHMGAQMVK.E
405 - 421	1567.9050	1566.8977	1566.8728	0.0249	0	R.AAVEEGVVAGGGVALIR.V
37 - 51	1589.8849	1588.8776	1588.8824	-0.0048	1	R.NVVLDKSFGAPTITK.D
43 - 58	1605.8633	1604.8560	1604.8522	0.0039	1	K.SFGAPTITKDGVSVAR.E
328 - 345	1845.9413	1844.9340	1844.9115	0.0225	0	K.DTTTIIDGVGEEAAIQGR.V
172 - 197	2851.3804	2850.3731	2850.3228	0.0503	0	K.EGVITVEDGTGLQDELDVVEGMQFDR.G

GroEL was removed in the presence of ATP at 37 ℃ but GST-S1 was degraded and a number of non-target proteins were detected (Figure 5.10 A). Removal of GroEL at a lower temperature was examined; however, both GST-S1 and GroEL were retained at 4 ℃ (Figure 5.10 B). At 25 ℃, most of the GST-S1 and GroEL were lost. In conclusion, addition of ATP may not be a good method for the removal of GroEL.

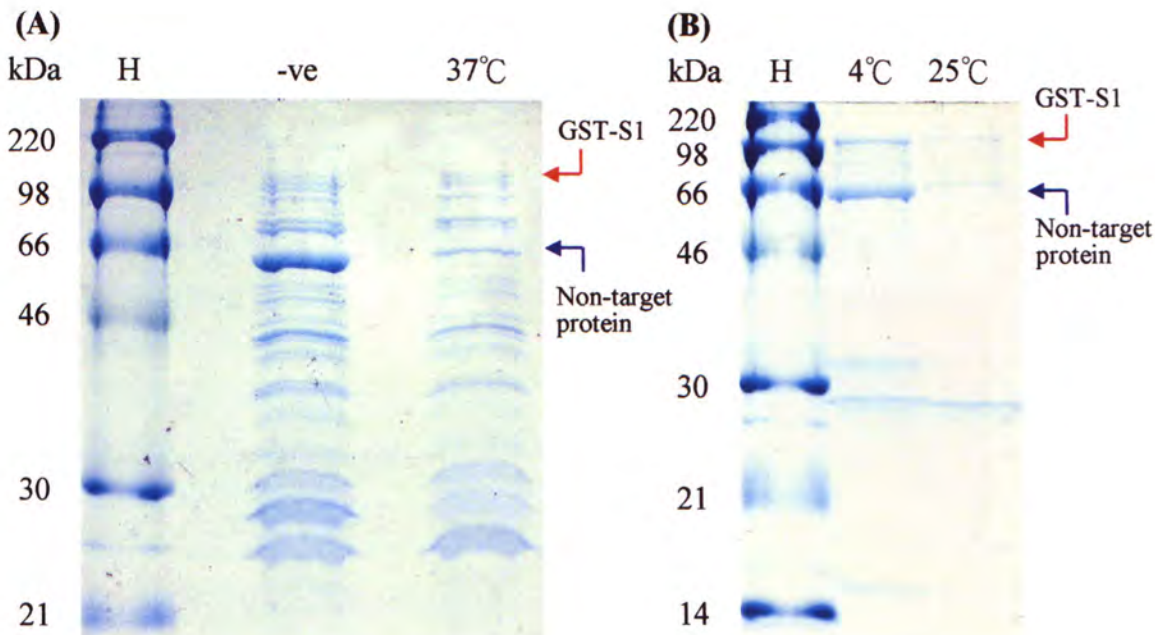


Figure 5.10 – Remove of GroEL by ATP in various temperatures. (A) GroEL was removed in presence of ATP but GST-S1 was lost as well. More non-target proteins were detected at 37 ℃. **(B)** Both of GST-S1 and GroEL was retained at 4 ℃ but most of them were lost at 25 ℃. H: High range marker. -ve: Without ATP at 37 ℃. Number: With ATP at temperatures. The red and blue arrows represent locations of GST-S1 and the 60 kDa non-target protein respectively.

On the other hand, GroEL was removed after washing by PBS with urea (Figure 5.11). Increase in the urea concentration of PBS caused loss of both GST-S1 and GroEL in eluents but the degree of the loss of GST-S1 was lower than that of GroEL. In PBS with 2M of urea, GST-S1 was still retained while GroEL was removed. However, other non-target proteins were still present in eluents. Thus, the sample was subjected to SEC. Elution profile and SDS-PAGE analysis showed that GST-S1 in the absence of GroEL was still eluted in the void volume, suggesting that GST-S1 formed aggregates or bound to other bacterial proteins (Figure 5.12 and 5.13).

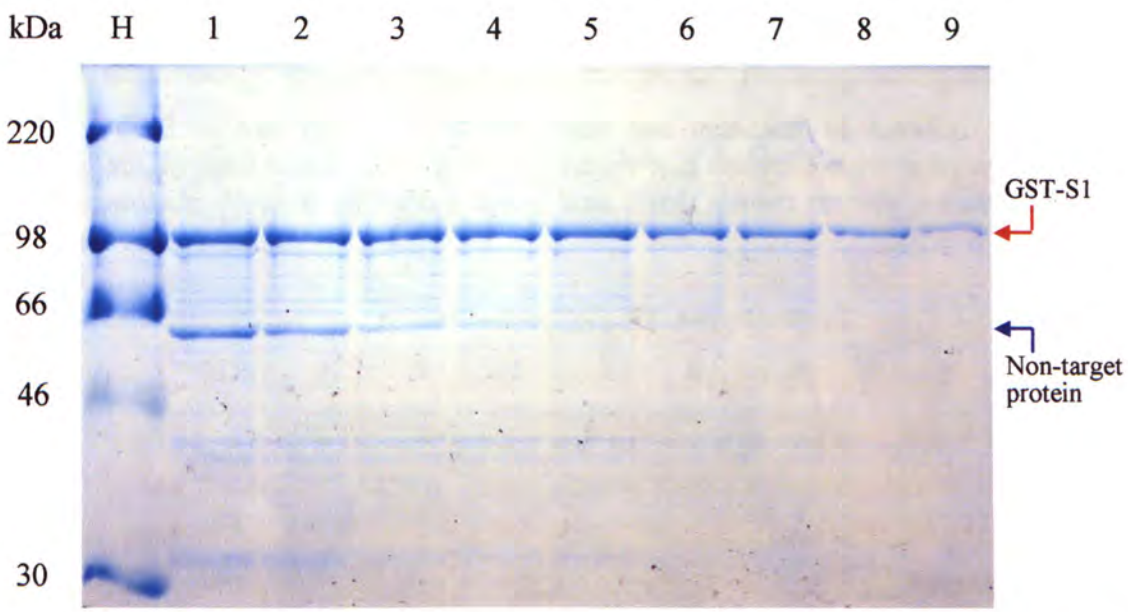


Figure 5.11 – Loss of GroEL in various urea concentrations. Both GST-S1 and GroEL in eluents were lost when the urea concentration of PBS was raised. In PBS with 2 M of urea, GST-S1 without loss had the highest purity. H: High range marker. 1 – 9: 0 M, 0.5 M, 1 M, 1.25 M, 1.5 M, 1.75 M, 2 M, 2.25 M and 2.5 M of urea. The red and blue arrows represent locations of GST-S1 and the 60 kDa non-target protein respectively.

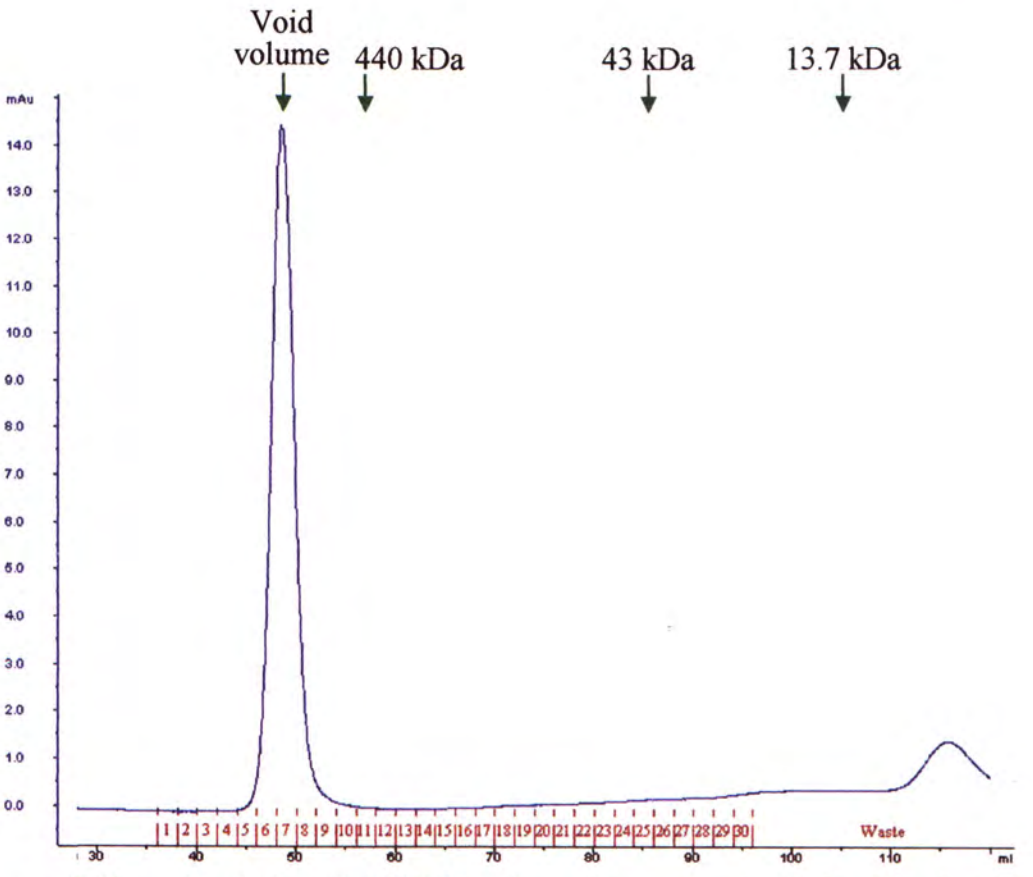


Figure 5.12 – Analysis of GST-S1 after the removal of GroEL by SEC. Only a single peak was eluted at 49 mL, which was the void volume by comparing with standards. Volume of each fraction was 2 mL. Green arrows indicate elution volumes of standards.

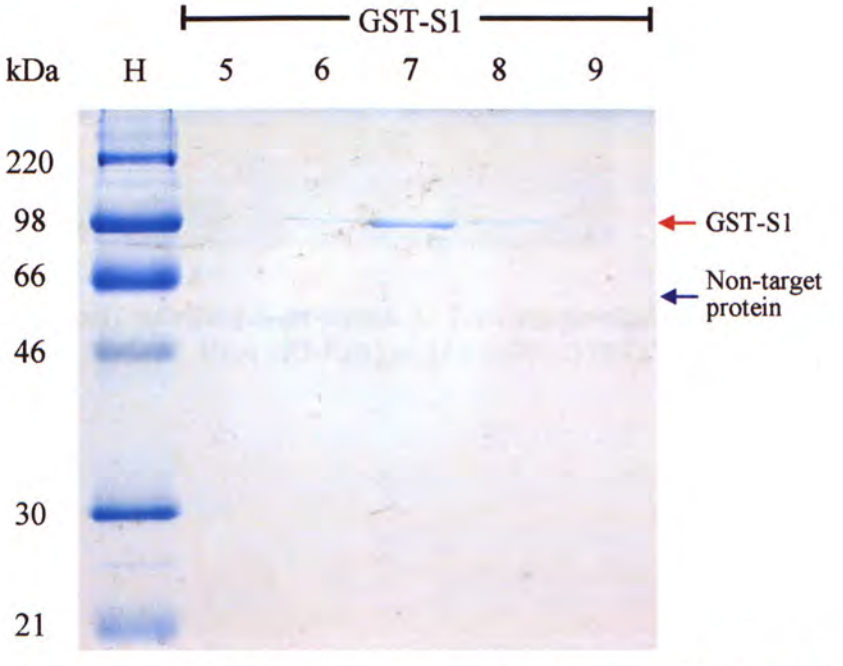


Figure 5.13 – GST-S1 without GroEL after purification by SEC. Both GST-S1 and non-target proteins were eluted in the void volume. H: High range marker. Number: Fraction number of SEC on figure 5.12. The red and blue arrows represent locations of GST-S1 and the 60 kDa non-target protein respectively.

5.3.4 Unfolding and Refolding

Bacterial inclusion bodies containing His-(15-317), His-(318-510), His-(587-826) and His-(903-1187) were solubilized in 6 M guHCl. These recombinant proteins were immobilized on His-binding resin, followed by washing and eluting by the buffer with 50 mM and 500mM of imidazole respectively. The purities of His-(567-826) and His-(903-1187) were approximately 90 % while His-(15-317) and His-(318-510) with small amounts of non-target proteins were present in eluents (Figure 5.14). Mass spectrometry analysis showed that all obvious non-target bands were degradative products except that the 45 kDa protein in eluent of His-(318-510) is an elongation factor.

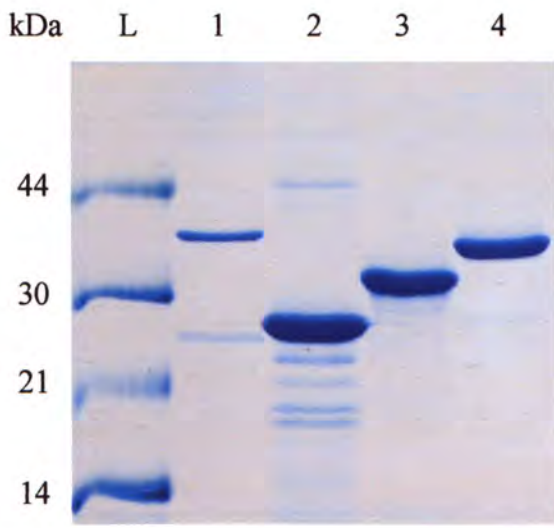
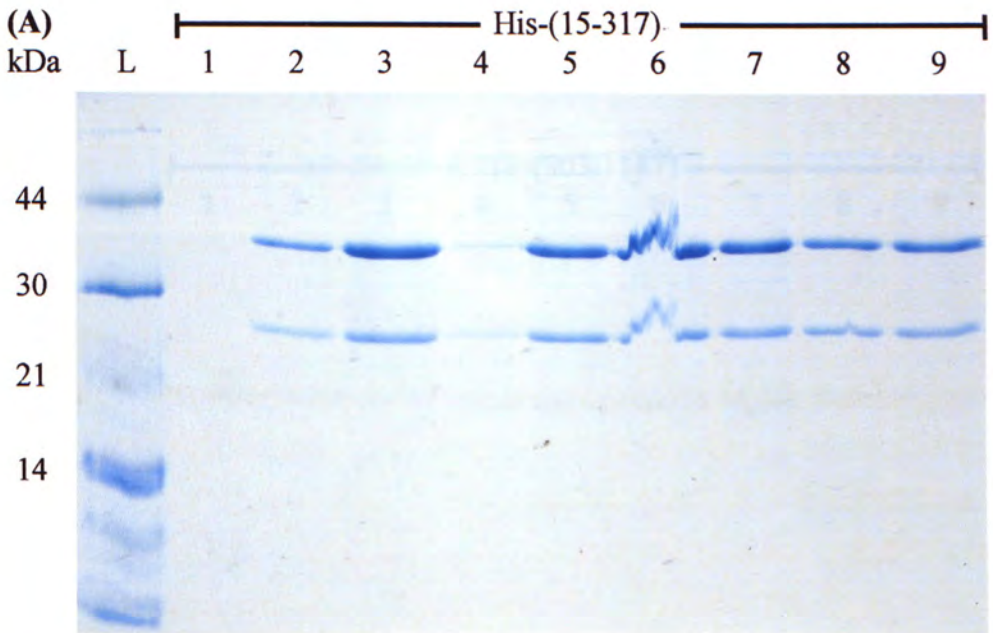
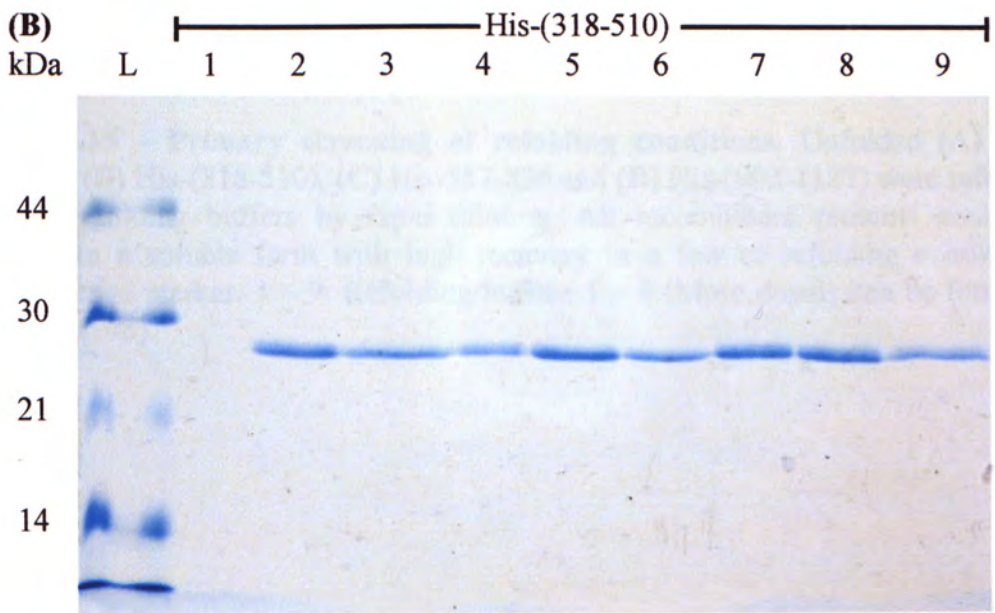


Figure 5.14 – Partially purified S proteins. L: Low range marker. 1 – 4: Eluents of His-(15-317), His-(318-510), His-(587-826) and His-(903-1187).

The partially purified proteins with final concentrations of 50 ng/μL were refolded in various refolding conditions by rapid dilution (Table 3.1). After centrifugation, supernatant was analyzed by SDS-PAGE to determine the recovery yields (Figure 5.15). Refolding buffer 2, one of the conditions with the highest yield, was selected for secondary and tertiary screening of refolding conditions of His-(318-510) (Table 3.1). We found that PEG and divalent cation did not have a significant effect on the improvement of protein solubility (Figure 5.16).





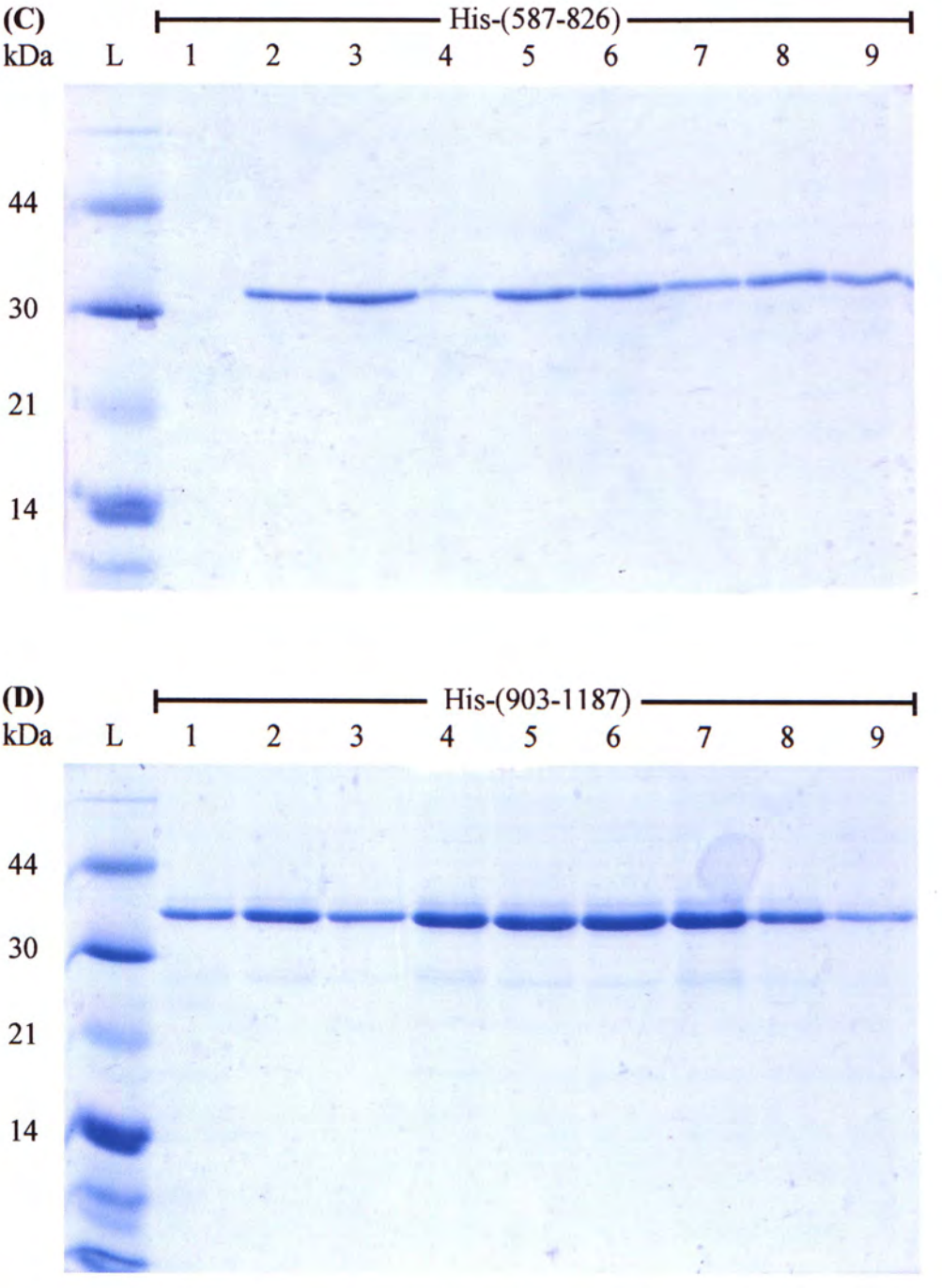


Figure 5.15 – Primary screening of refolding conditions. Unfolded **(A)** His-(15-317), **(B)** His-(318-510), **(C)** His-587-826 and **(D)** His-(903-1187) were refolded in nine refolding buffers by rapid dilution. All recombinant proteins could be refolded in a soluble form with high recovery in a few of refolding conditions. L: Low range marker. 1 – 9: Refolding buffers 1 – 9 (More details can be found in table 3.1, p.45).

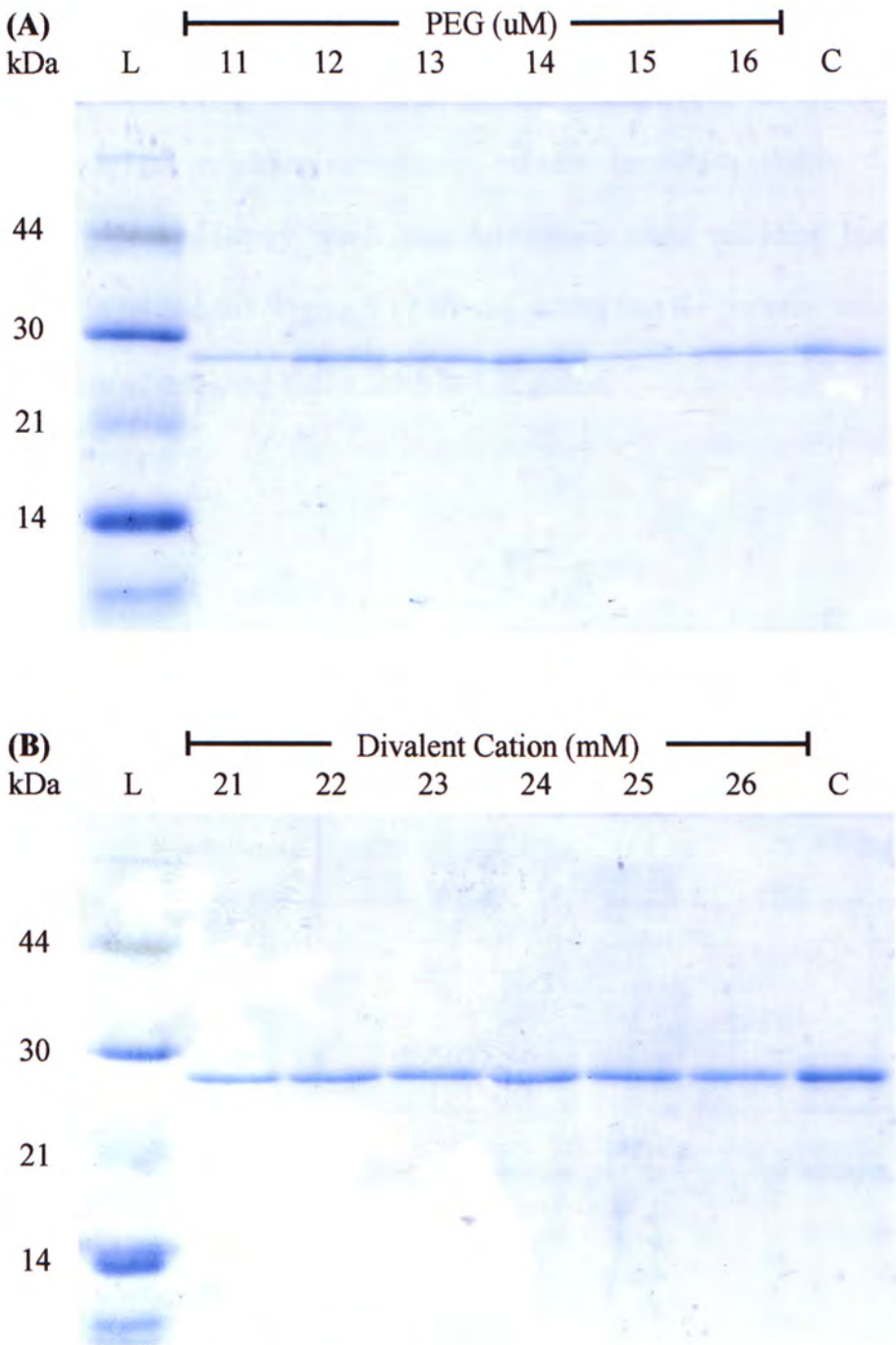


Figure 5.16 – Secondary and tertiary screening of refolding conditions of His-(318-510). Refolding conditions of His-(15-317) were **(A)** secondary and **(B)** tertiary screened by rapid dilution. The amounts of soluble recombinant proteins refolded in various conditions were similar and that recovery rates closed to 100 %. L: Low range marker. 11 – 16: Refolding buffers 1 – 6 of secondary screening. 21 – 27: Refolding buffers 1 – 6 of tertiary screening. C: Total protein.

To further confirm whether the protein was refolded successfully, refolded His-(318-510) was rapidly diluted in PBS. Small proportion of His-(318-510) refolded in various refolding conditions became insoluble (Figure 5.17 A). Nevertheless, a contradictory result was determined when refolding buffer was slowly removed by dialysis (Figure 5.17 B), suggesting that the proteins were soluble only in presence of refolding buffer, such as L-arginine.

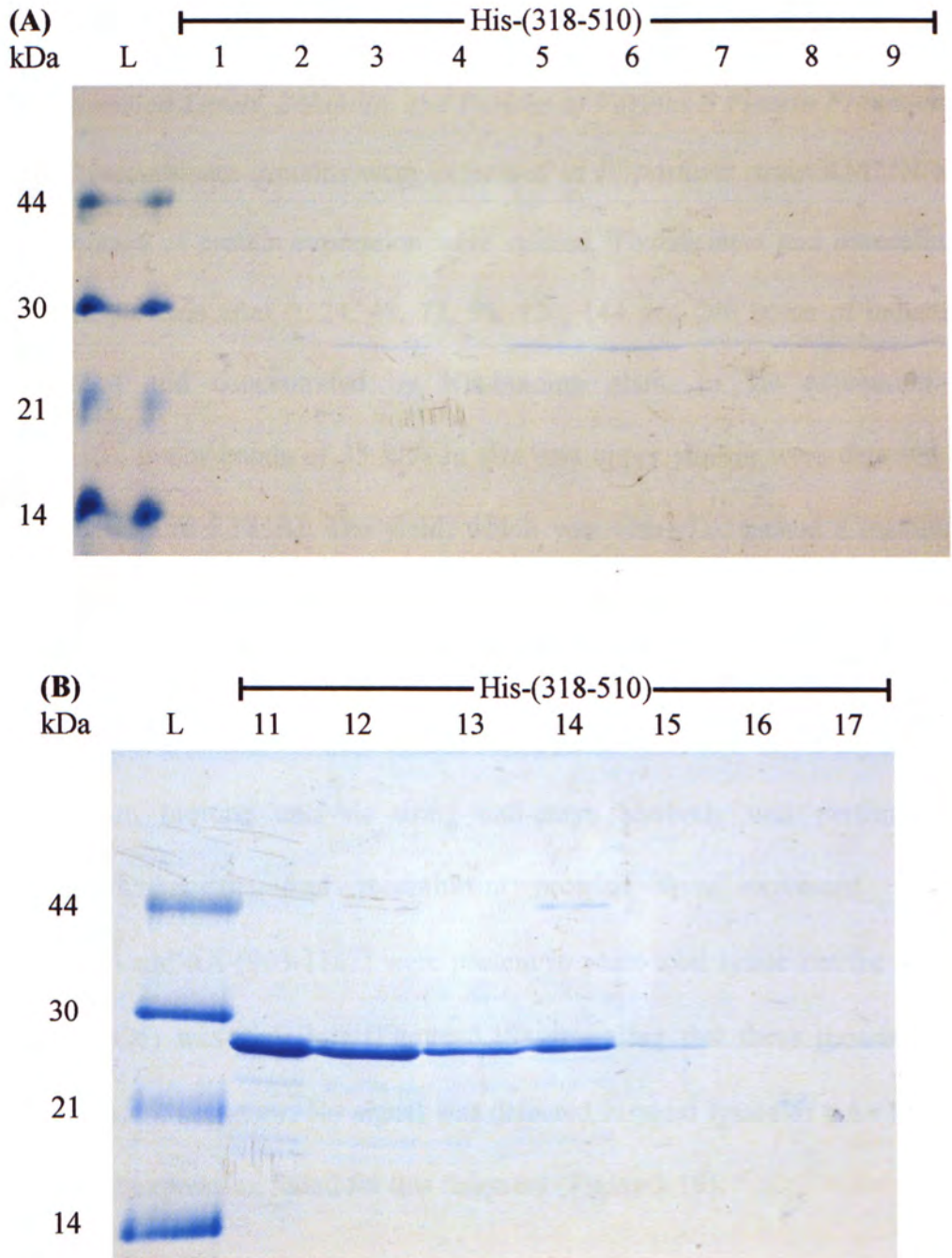


Figure 5.17 – Solubility of refolded His-(318-510) in PBS. (A) Little soluble His-(318-510) refolded in various conditions became insoluble only after rapidly diluting in PBS. **(B)** The soluble His-(318-510) amount was depleted after dialysis by PBS. L: Low range marker. 1 – 9: Insoluble fraction of refolded protein after rapid dilution by PBS. 11: Total protein. 12: Refolded total protein. 13: Refolded soluble protein. 14: Dialyzed total protein. 15: Dialyzed soluble protein. 16: Flow through. 17: Eluent.

5.4 Protein Expression and Optimization in *P. pastoris*

5.4.1 *Expression Levels, Solubility and Purities of Various S Protein Fragments*

Six recombinant proteins were expressed in *P. pastoris* strain KM71H and the time courses of protein expression were studied. The secreted and intracellular recombinant proteins after 0, 24, 48, 72, 96, 120, 144 and 240 hours of induction were purified and concentrated by His-binding resin. In the expression of αA-(318-510), major bands of 35 kDa in size and upper smears were detected by SDS-PAGE (Figure 5.18 A). The yield, which was 46mg/L, reached a maximum after 144 hours of induction. However, detection of other recombinant proteins by SDS-PAGE failed except αA-(15-317) (Figure 5.18 B). The yield reached a maximum after 144 hours of induction as well but the yield was only 5 mg/L.

Western blotting analysis using anti-cmyc antibody was performed to determine whether the four recombinant proteins were expressed. αA-S2, αA-(587-826) and αA-(903-1187) were present in yeast total lysate but the amount of αA-(587-826) was very low (Figure 5.19), revealing that these proteins were expressed but not secretory. No signal was detected in yeast lysate of αA-(13-672), indicating that expression failed for this fragment (Figure 5.19).

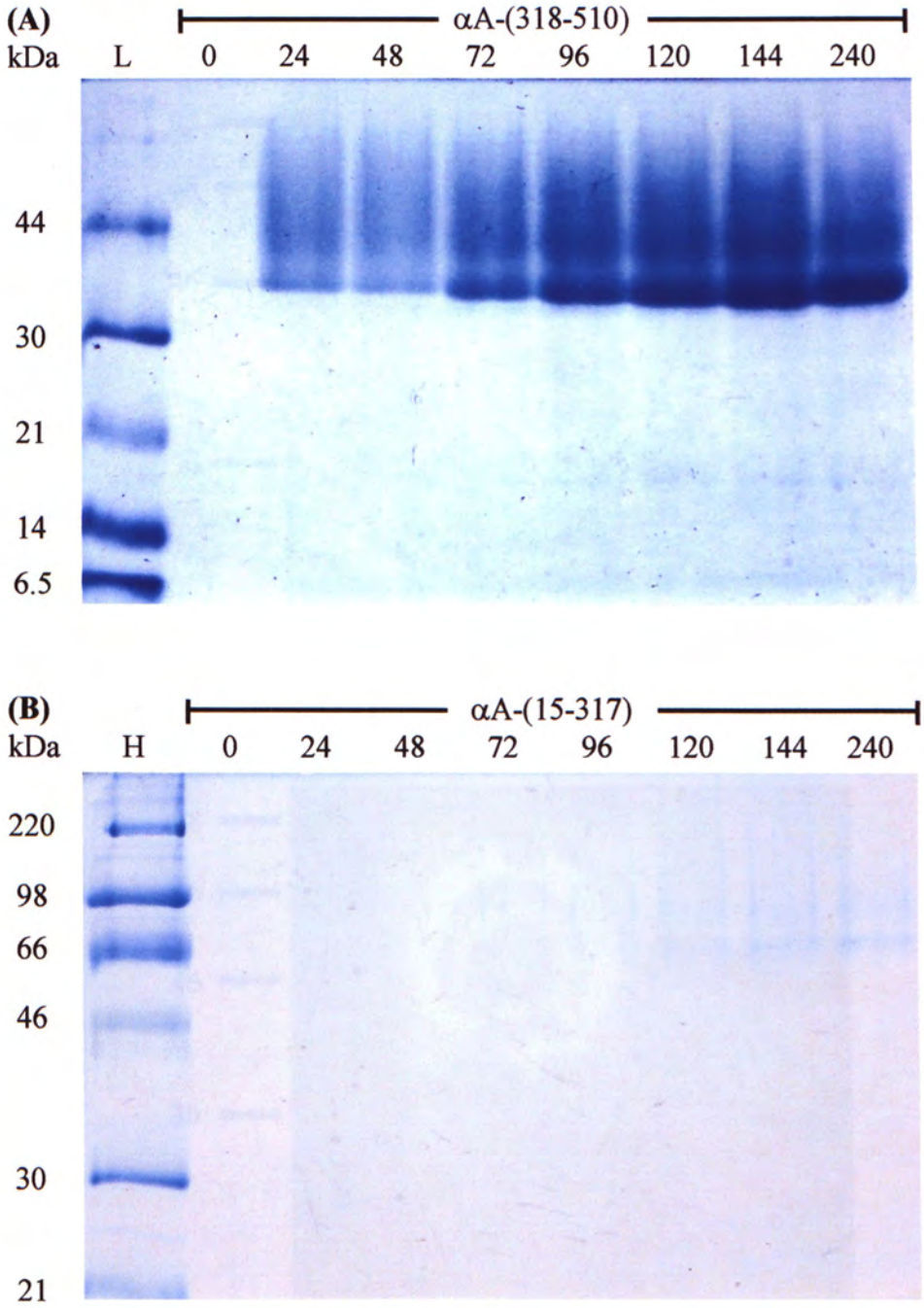


Figure 5.18 – Time course study of secretory S protein fragments. **(A)** The 35 kDa bands and upper smears were detected after 24 hours of induction of αA-(318-510) expression. **(B)** Smears of αA-(15-317) were detected.

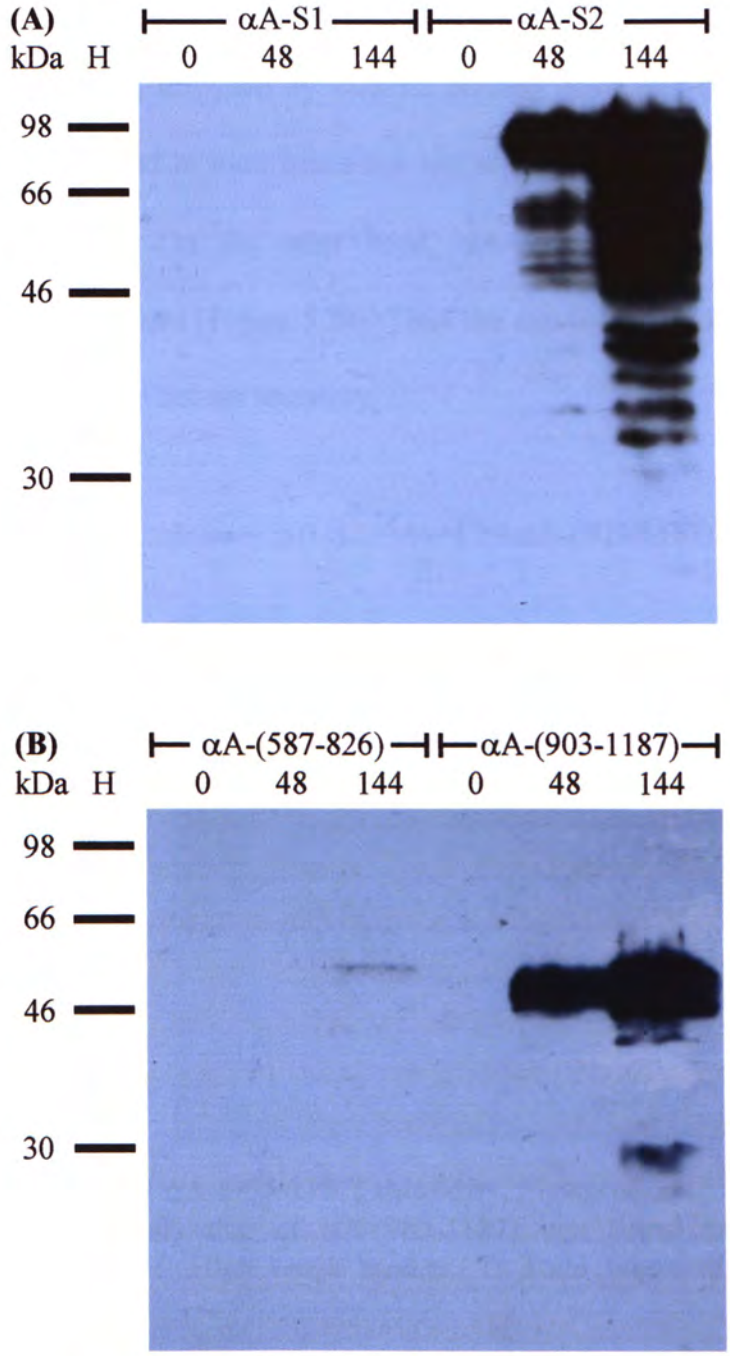


Figure 5.19 – Presence of intracellular S protein fragments in *P. pastoris*. **(A)** Signals of αA-S2 but not αA-S1 were found in yeast lysate after induction of 144 hours. **(B)** Both αA-(587-826) and αA-(903-1187) were present in the lysate but the αA-(587-826) amount was much lower. H: High range marker. Numbers: yeast lysate after hours of induction.

αA-S2 and αA-(903-1187) in the yeast lysate were purified by affinity chromatography and then analyzed by western blotting using anti-cmyc antibody. αA-S2 was only presented in total lysate but not supernatant, showing that it was insoluble (Figure 5.20). On the other hand, αA-(903-1187) was detected in supernatant as well as eluent (Figure 5.20). Thus the expression of this fragment is successful and it is soluble but not secretory.

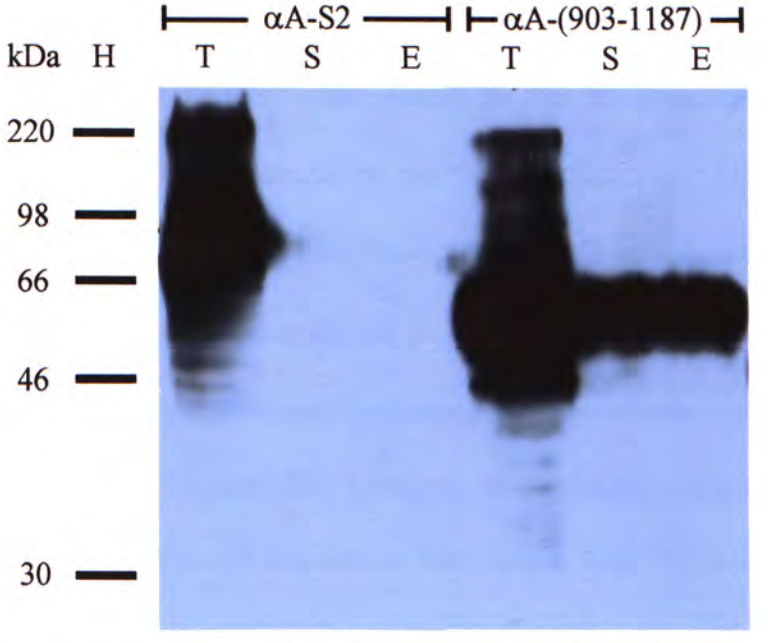


Figure 5.20 – αA-S2 and αA-(903-1187) solubility. Signal of αA-S2 was detected in yeast lysate only while that of αA-(903-1187) was found in total lysate, supernatant and eluent. H: High range marker. T: Total lysate. S: Supernatant. E: Eluent.

5.4.2 *Characterization of De-N-glycosylated Recombinant Proteins*

All sizes of major bands and smears were higher than the expected size. One of the possible reasons is glycosylations of the recombinant proteins since proteins expressed in yeast are commonly highly-glycosylated. If it is true, sizes of the recombinant proteins after deglycosylation should decrease to the expected size. SDS-PAGE analysis showed that αA-(318-510) de-N-glycosylated by PNGase F was 29 kDa in size and a few proteins larger than 29kDa were presented (Figure 5.21 A). Their sizes were still larger than 24 kDa, the calculated size of αA-(318-510). The increased size might be due to O-glycosylations or incomplete de-N-glycosylation. Results of de-N-glycosylated αA-(15-317) and αA-(903-1187) were similar to that of αA-(318-510) (Figure 5.21 B and C).

The possibility of the presence of α signal peptide in αA-(318-510) N-terminal was ruled out since N-terminal sequencing results showed that the signal peptide was cleaved by Kex2 but not Ste13, leading to the presence of four extra a. a. residues (Figure 5.22). Failure of translation termination was impossible as well because peptide mass fingerprint had a peak of 2126.9001 Da, which matched the calculated mass of C-terminal peptide, 2126.9750 Da (Figure 5.23).

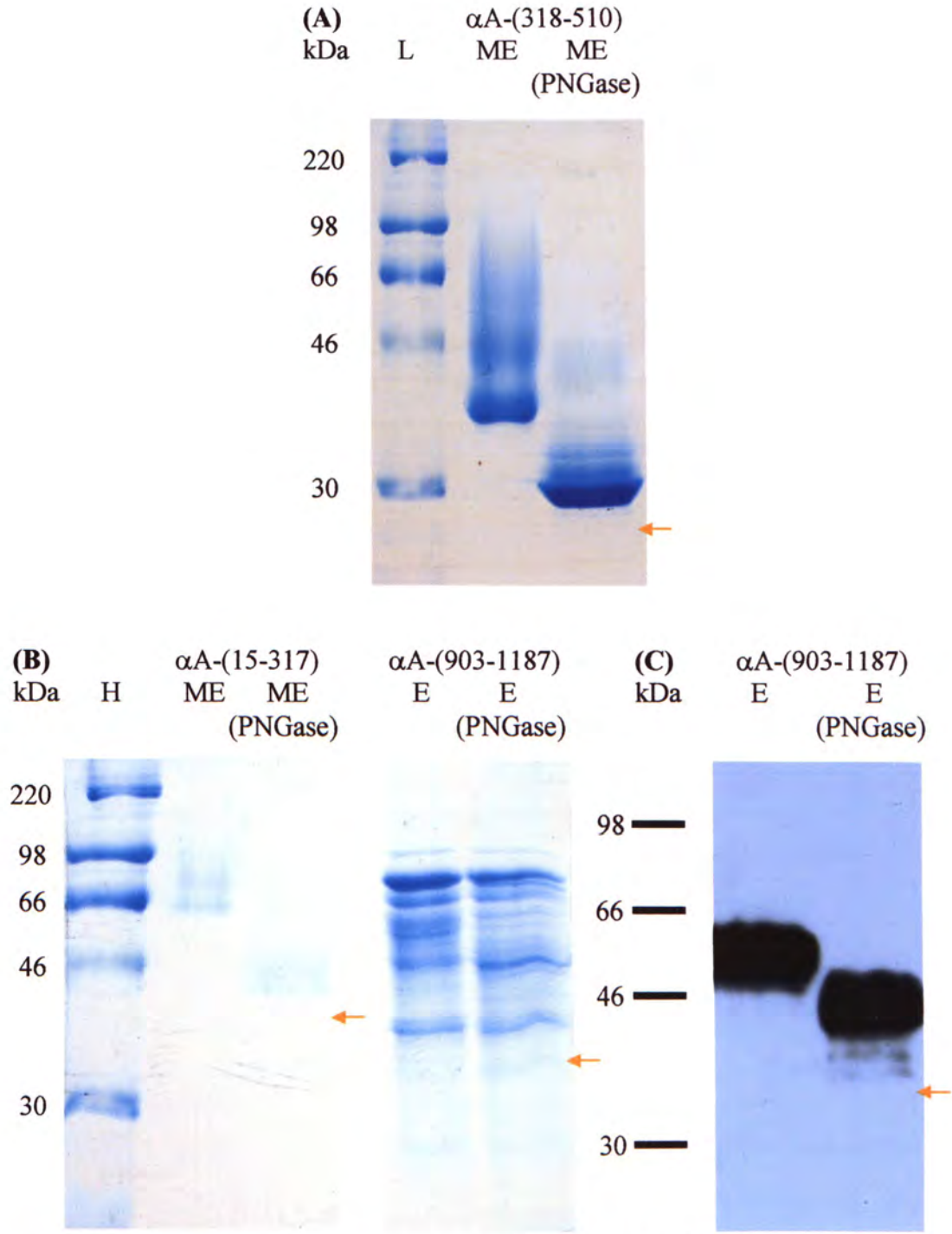


Figure 5.21 – De-N-glycosylation of αA-(318-510), αA-(15-317) and αA-(903-1187) by PNGase F. (A) Both major band and smear of αA-(318-510) were decreased in sizes. The major band intensity was increased while the smear intensity was decrease. **(B, C)** The sizes of smears of αA-(15-317) and αA-(903-1187) were decreased while their intensities were unchanged. L: Low range marker. H: High range marker. ME: Eluent of medium. E: Eluent of yeast lysate. (PNGase): De-N-glycosylation by PNGase F. Arrows indicate the expected band locations.

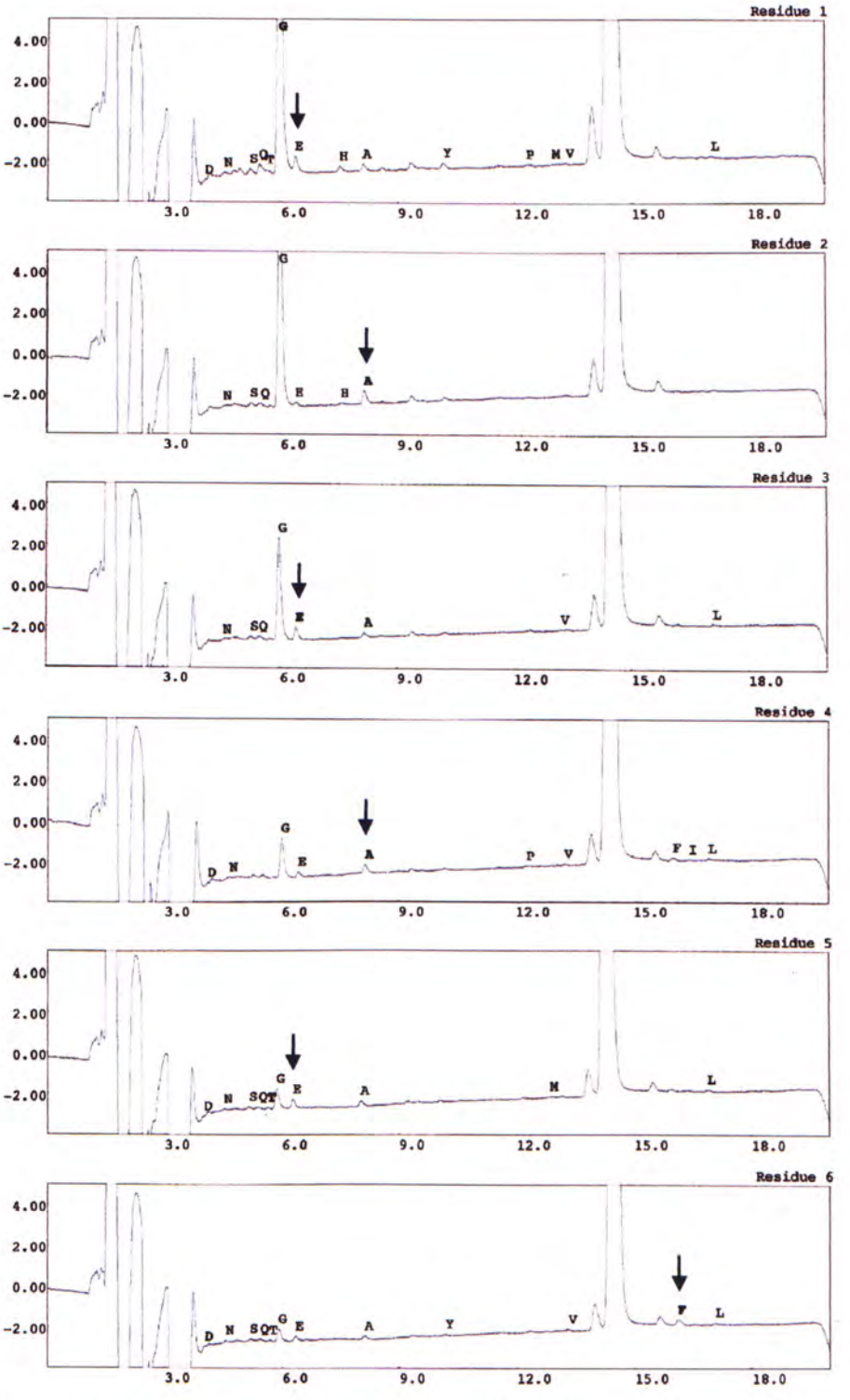


Figure 5.22 – N-terminal sequencing of αA-(318-510). The N-terminal sequence of αA-(318-510) was EAEAEF, which was matched to the N-terminal sequence of the recombinant protein cleaved by Kex2 but not Ste13. Blue arrows indicate the a. a. residue 1 – 6.

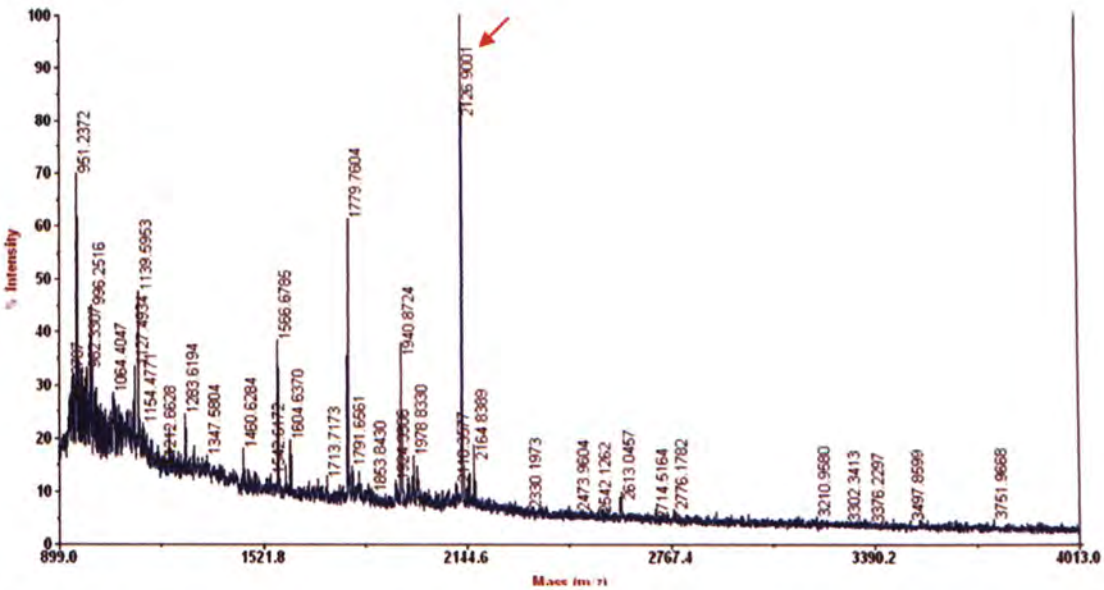


Figure 5.23 – PMF of αA-(318-510). The peak of 2126.9001 Da matched C-terminal of αA-(318-510), LISEEDLNSAVDHHHHHH, with a calculated mass of 2126.9750 Da. The red arrow indicates the matched peak.

The GST-(318-510) expressed in *E. coli* formed aggregates while the native conformation of RBD was monomer. The native size of αA-(318-510) was determined by SEC. A board peak was eluted between 58 mL and 88 mL while the highest reading was at 78 mL (Figure 5.24). The eluents were analyzed by SDS-PAGE and results showed that fraction 25 contained only a band of αA-(318-510) but not a smear (Figure 5.25). By comparing with standards of SEC, native size of the protein eluted in fraction 25 was 34 kDa, which matched the molecular weight of αA-(318-510) determined by SDS-PAGE. In conclusion, αA-(318-510) was expressed in a monomeric form. The difference in the sizes of proteins in earlier fractions determined by SEC and SDS-PAGE might be because of glycosylation.

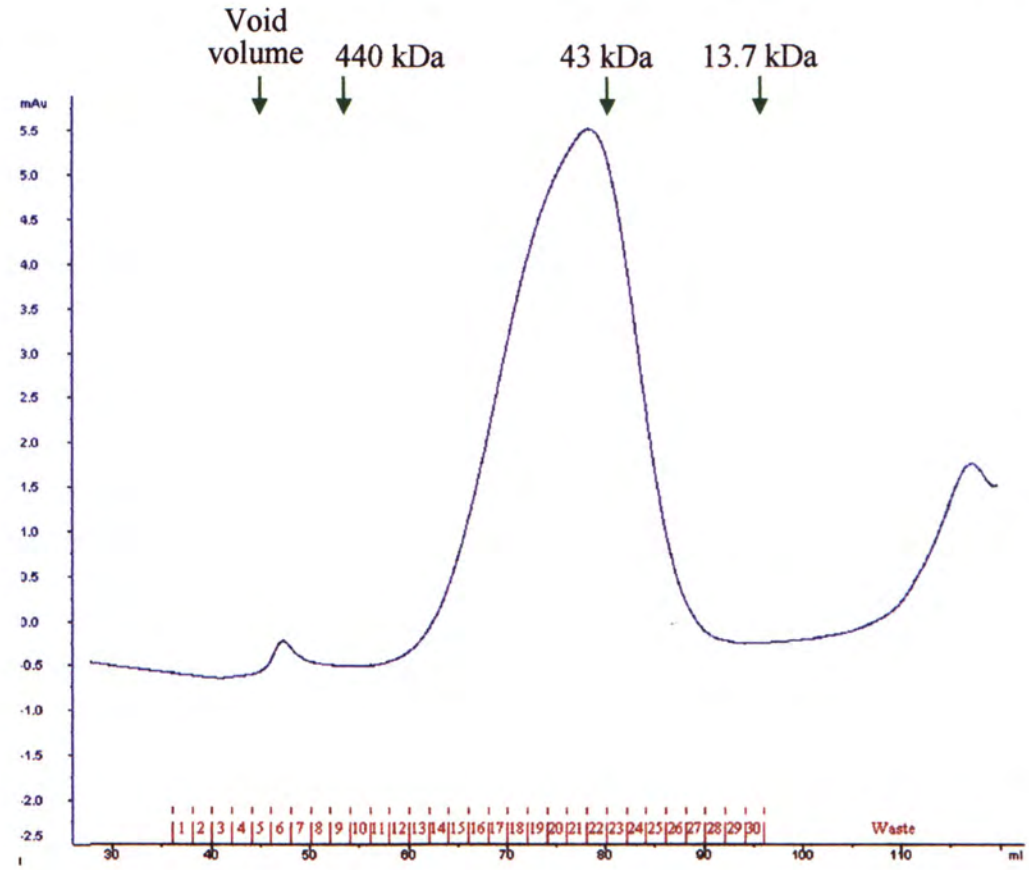


Figure 5.24 – Determination of αA-(318-510) native size by SEC. A board peak was eluted between 52 mL and 88 mL and the peak height was at 78 mL, which was 60 kDa by comparing with standards. Volume of each fraction was 2 mL. Green arrows indicate elution volumes of standards.

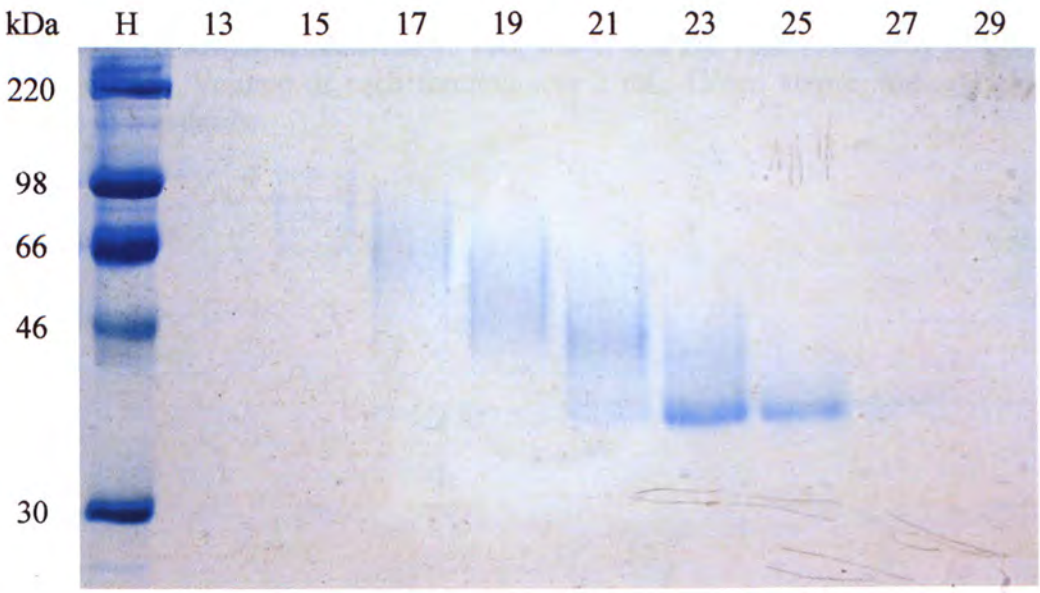


Figure 5.25 – SDS-PAGE of αA-(318-510) purified by SEC. H: High range marker. Number: Fraction number of SEC on figure 5.24.

De-N-glycoslyated αA-(318-510) was examined in order to further confirm its native size. The protein was nonetheless eluted in void volume, indicating that de-N-glycosylated αA-(318-510) was aggregated (Figure 5.26).

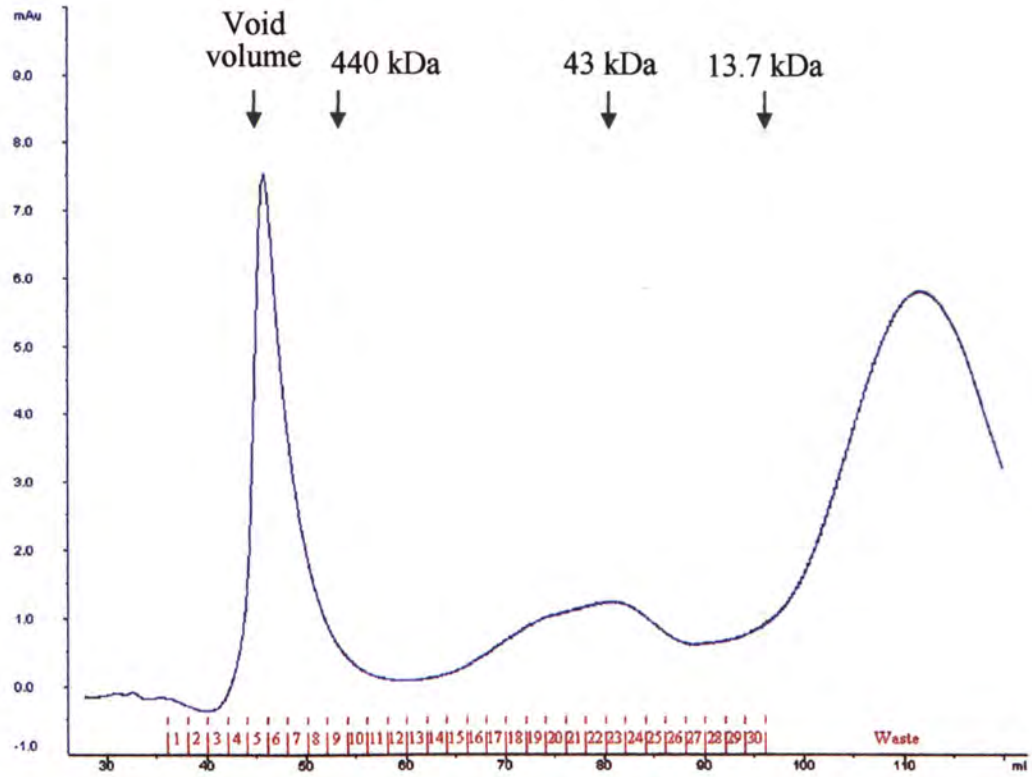


Figure 5.26 – Determination of de-N-glycoslyated αA-(318-510) native size by SEC. αA-(318-510) was eluted at 45 mL, which was the void volume by comparing with standards. Volume of each fraction was 2 mL. Green arrows indicate elution volumes of standards.

Identification of Interacting Partners

6.1 Practicability of Pull-down Assay

6.1.1 ACE2 Extraction

ACE2, a function receptor of SARS-CoV, binds to a. a. residue 318 – 510 of the S protein (Li *et al.*, 2003; Wong *et al.*, 2004). Since this protein is a TM protein located on Vero E6 cell surface, the presence of hydrophobic regions may deplete ACE2 so that it cannot be extracted. Confirmation of the presence of ACE2 after extraction by mammalian cell lysis buffer was therefore performed. Result of western blotting analysis using anti-ACE2 antibody revealed that ACE2 was present in soluble fractions of mammalian cell lysis buffer (Figure 6.1). On the other hand, less ACE2 present in the total lysate might be because of experimental error during transfer of Vero E6 proteins.

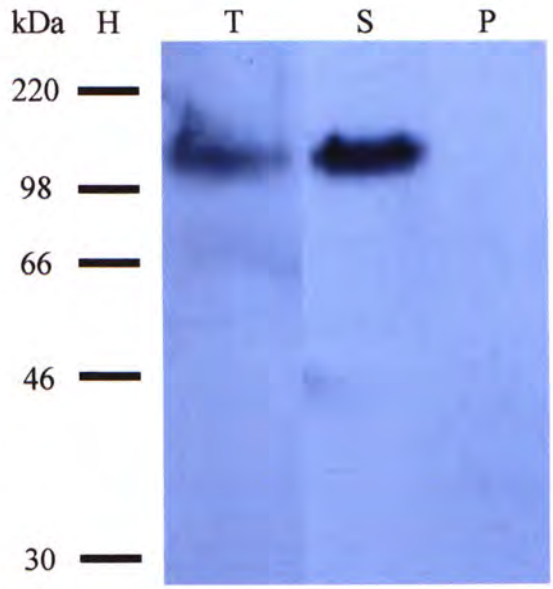
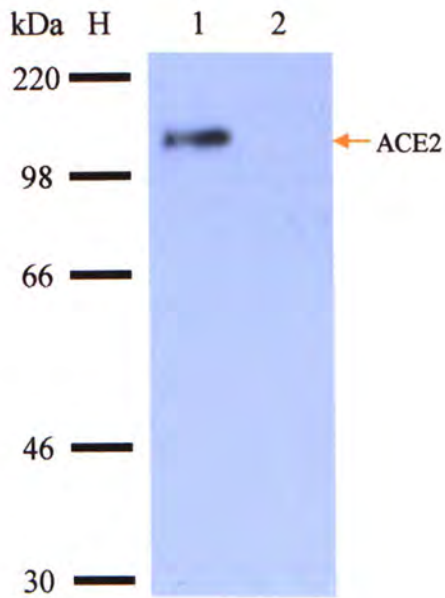


Figure 6.1 – The detection of ACE2 in the protein extracts. ACE2 of approximately 120 kDa was detected in the soluble fraction of mammalian cell lysis buffer but not insoluble fraction. H: High range marker. T: Total lysate. S: Soluble fraction. P: Insoluble fraction.

6.1.2 Pull-down of ACE2 by the P. pastoris-expressed recombinant RBD

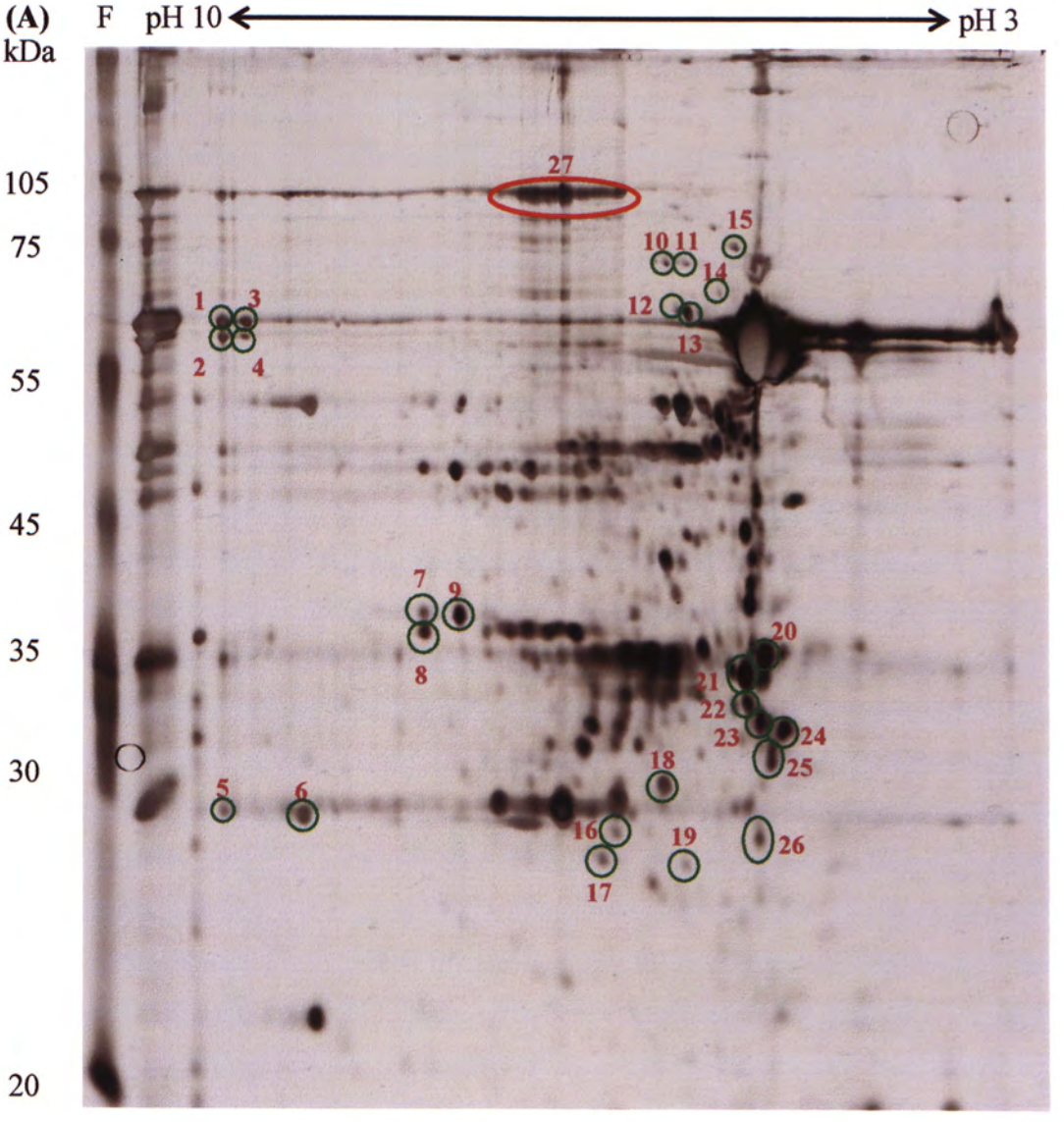
To confirm whether the expressed S protein was functional, pull-down assay using αA-(318-510) and Vero E6 cell lysate was performed. Presence of ACE2 in the eluate was then detected by western blotting using anti-ACE2 antibody. The western blotting result showed that the band representing ACE2 could be detected in the eluate of αA-(318-510) but not the negative control, in which αA-(318-510) was absent (Figure 6.2). The specific interaction between ACE2 and αA-(318-510) implied αA-(318-510) was functional and ACE2 as well as other interacting partners could be captured by pull-down assay.

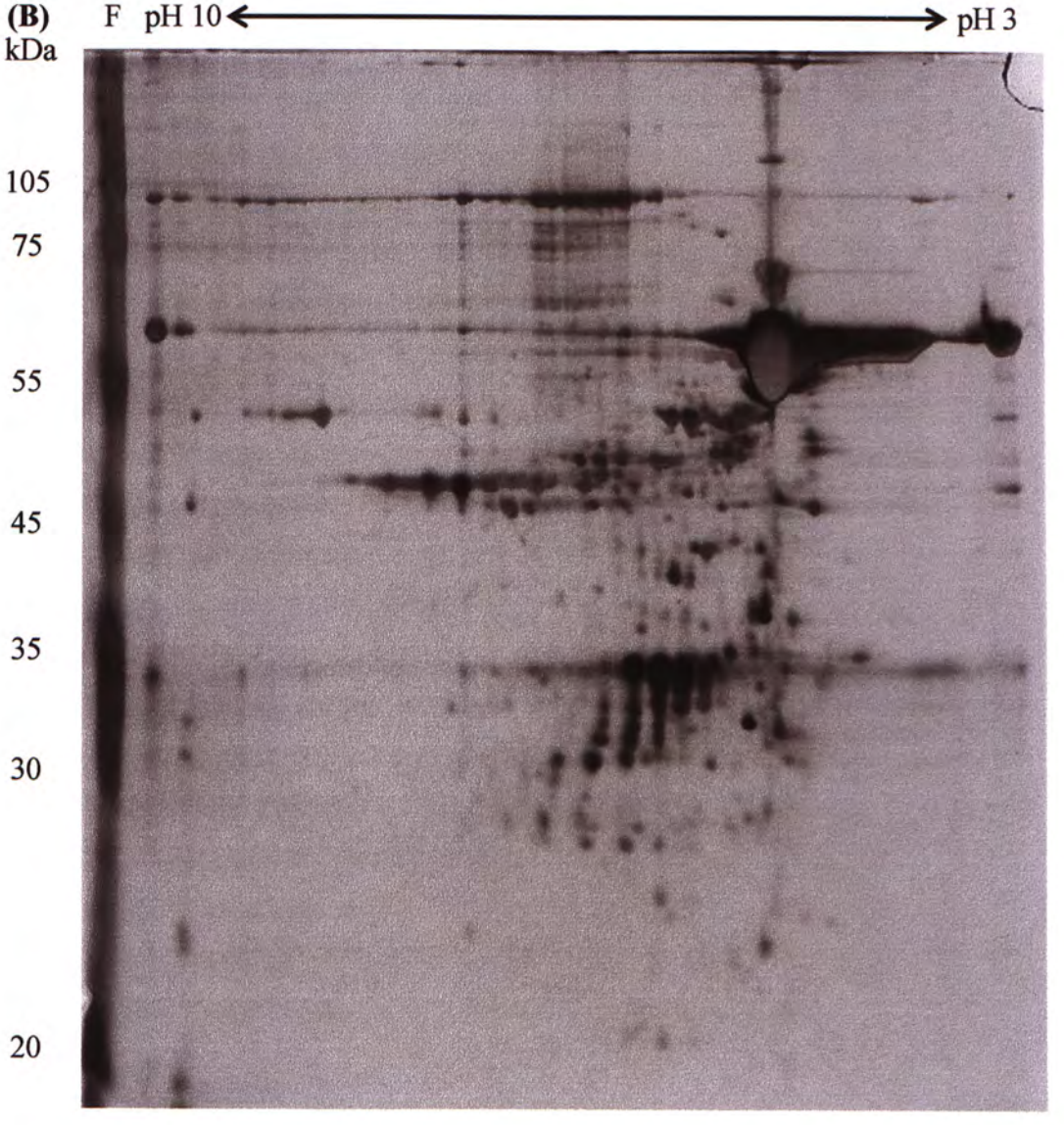
Figure 6.2 – Identification of ACE2 by pull-down assay using αA-(318-510). H: High range marker. 1: Presence of αA-(318-510). 2: Absence of αA-(318-510) as negative control. Orange arrow indicates the expected location of ACE2.

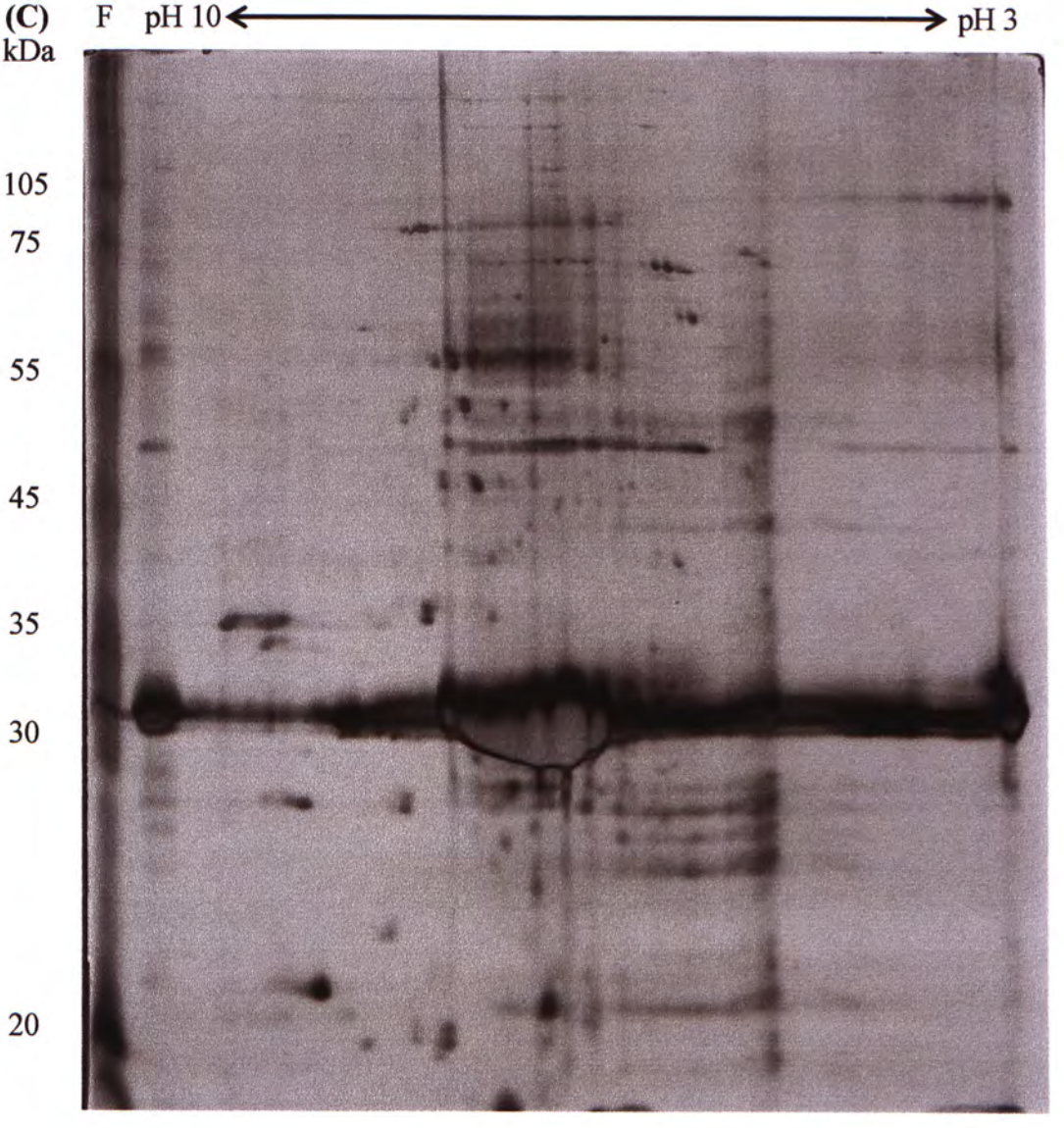


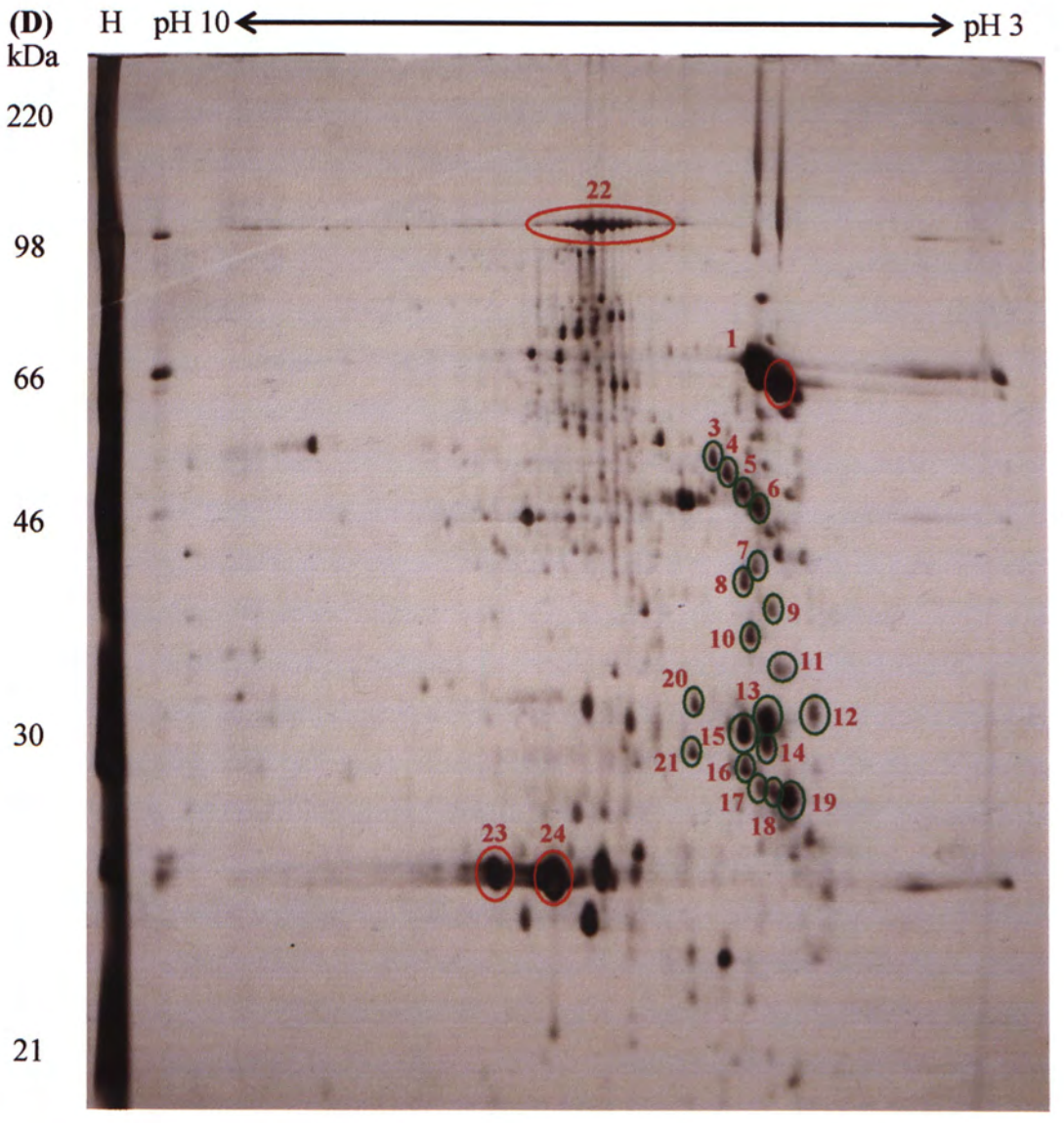
6.2 Pull-down Assay and Two-dimensional Gel Electrophoresis

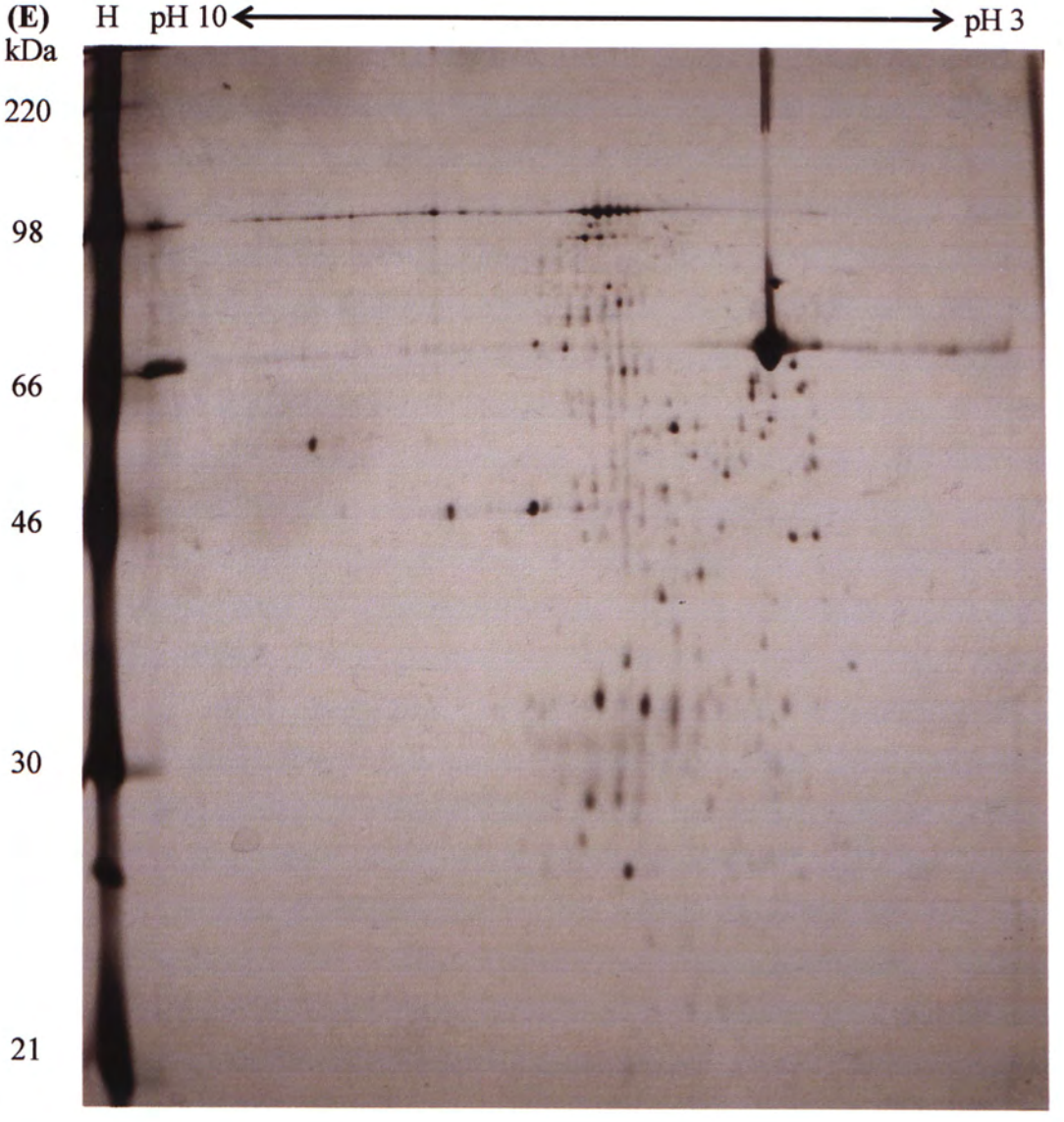
By comparing spots on Two-dimensional (2D) gels, approximately twenty to thirty spots were detectable only in the 2D gel using GST-S1 and Vero E6 cell lysate (Figure 6.3 and 6.4). A few of these spots were still detectable when pull-down assay using GST-S1 with GroEL and Vero E6 cell lysate was repeated. Interestingly, when we examined the results from GST-S1 purified by 2M of urea and Vero E6 cell lysate, locations of putative interacting partners were significantly changed. It might be due to the depletion of GroEL which changed the GST-S1 structure, or altered the amounts and conformations of other bacterial proteins. Thus, Vero E6 proteins interacting with them were different.


(A)
kDa
F
pH 10
pH 3
105
75
55
45
35
30
20
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27


(B)
kDa
F
pH 10
pH 3
105
75
55
45
35
30
20


(C)
F
pH 10
pH 3
kDa
105
75
55
45
35
30
20


(D)
H
pH 10
pH 3
kDa
220
98
66
46
30
21
22
1
3
4
5
6
7
8
9
10
11
20
13
12
15
14
21
16
17
19
18
23
24


(E)
H
pH 10
pH 3
kDa
220
98
66
46
30
21

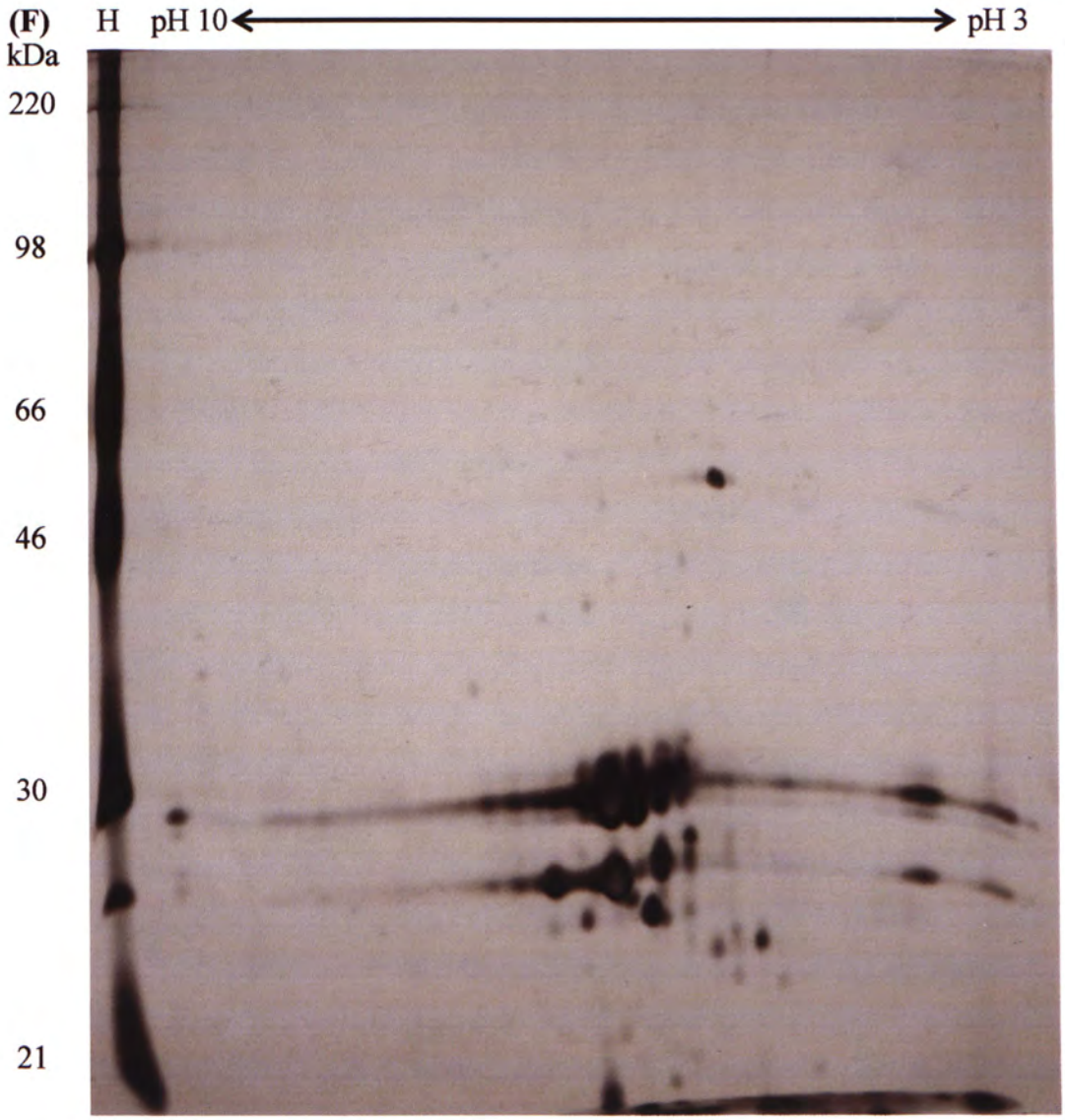
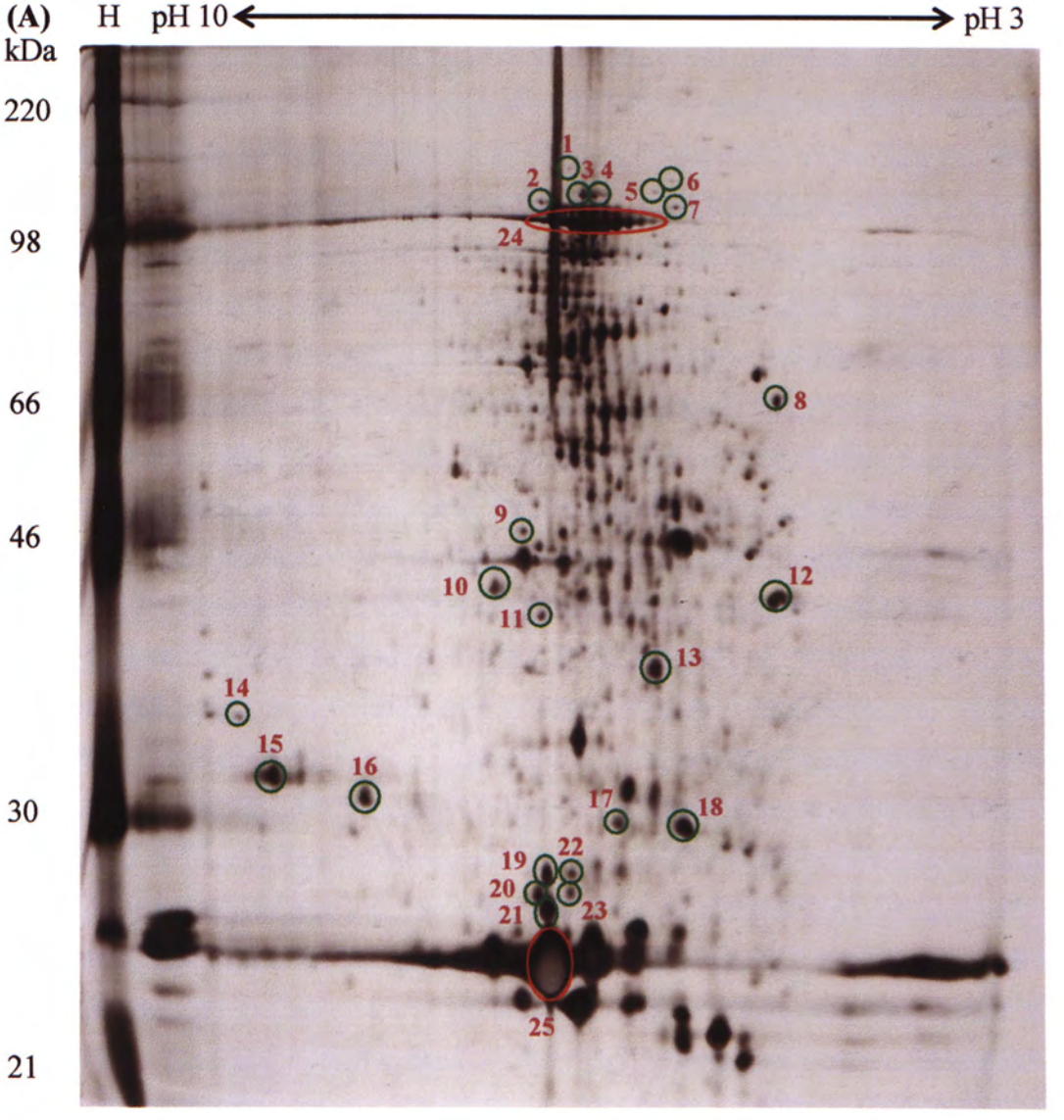
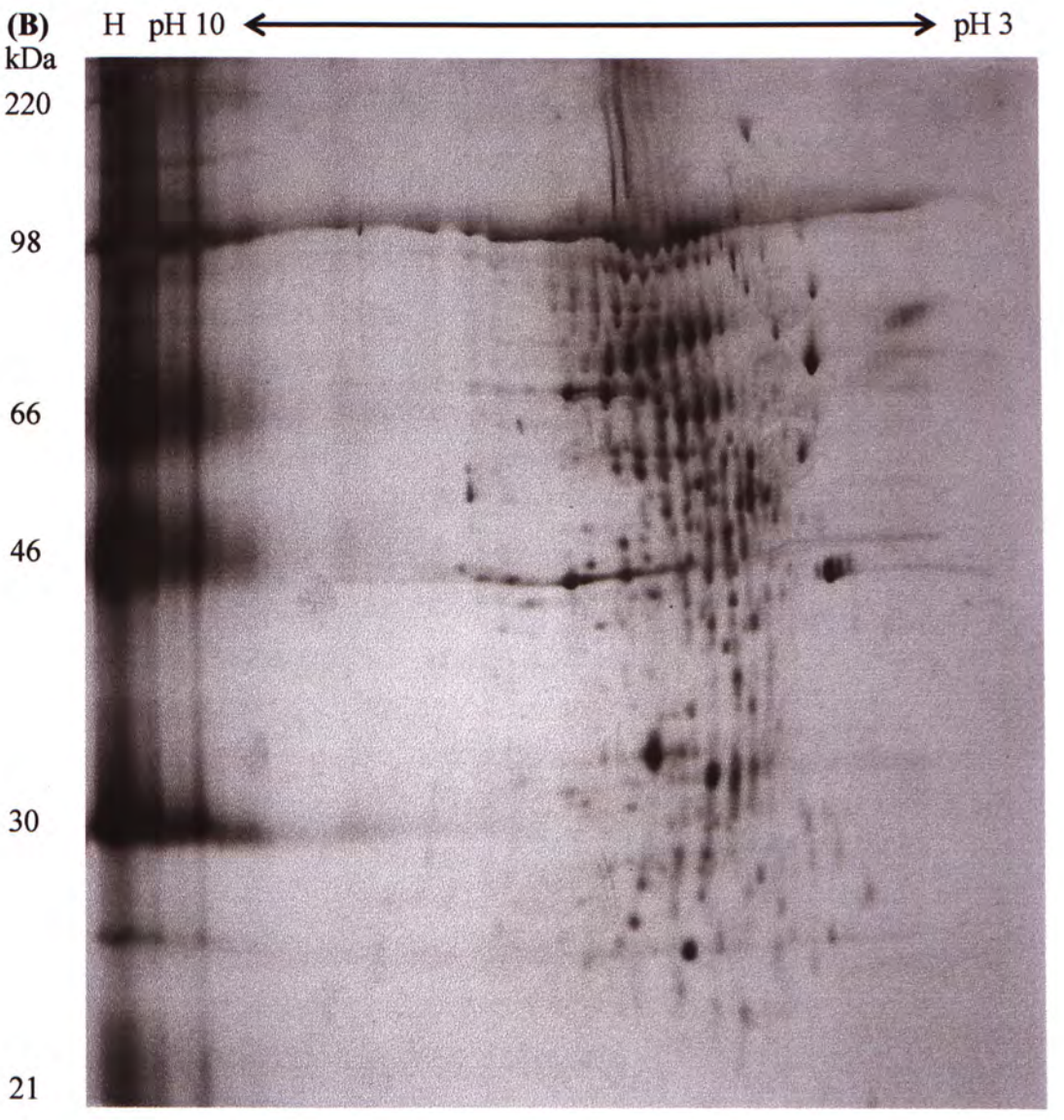


Figure 6.3 – Protein 2D map of GST-S1 with GroEL interacting with Vero E6 cell lysate. (A) GST-S1 with Vero E6 cell lysate, **(B)** GST-S1 without lysate and **(C)** GST with lyate were used for one of the trials of pull-down assay. **(D-F)** A few spots could be found in another trial of pull-down assay. Red circles indicate GST-S1 and GST while green circles represent spots that are detectable in the gel using GST-S1 and Vero E6 lysate. F: Full range marker.


(A)
H
pH 10
pH 3
kDa
220
98
66
46
30
21
1
2
3
4
5
6
7
24
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
25


(B)
H pH 10
pH 3
kDa
220
98
66
46
30
21

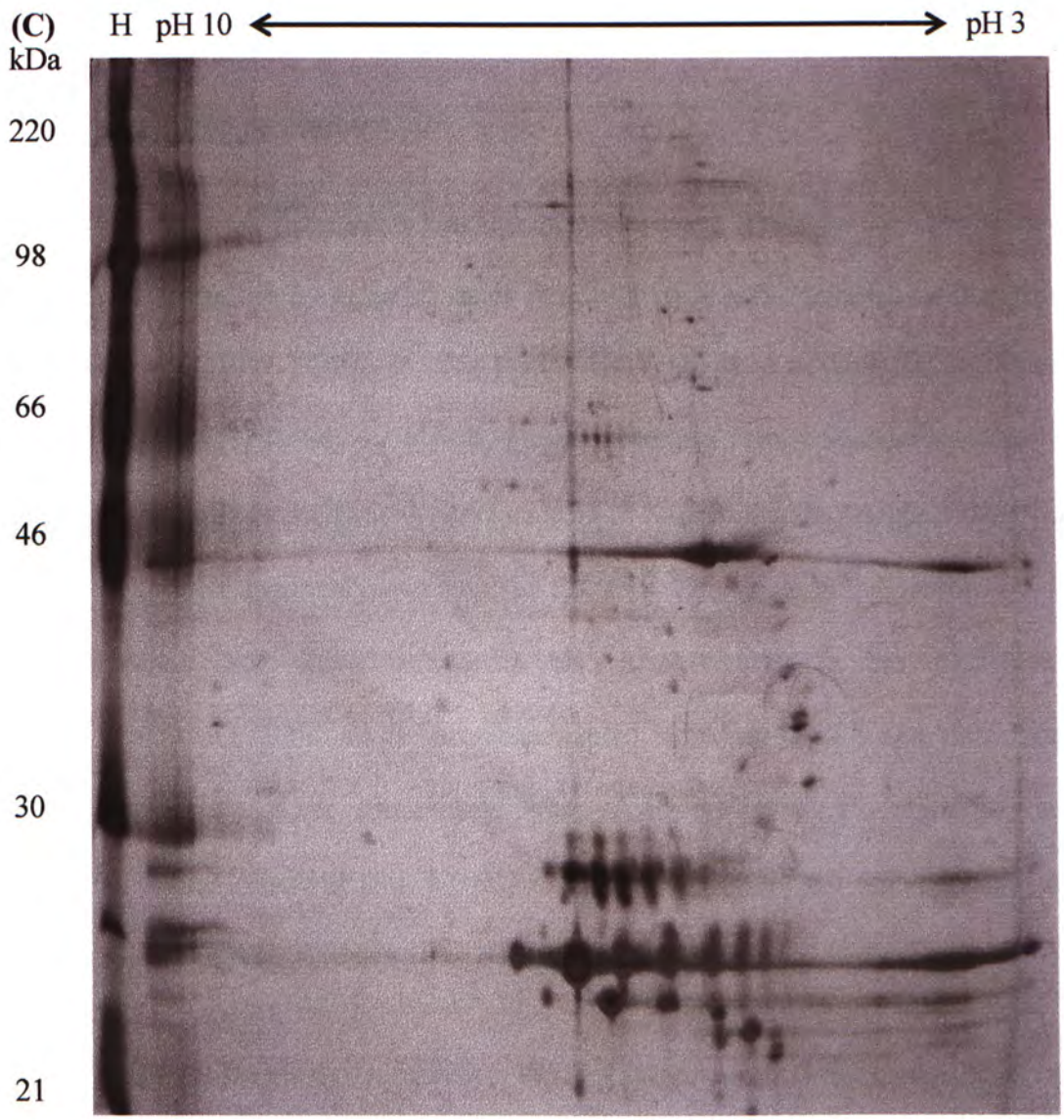


Figure 6.4 – 2D map of proteins generated by GST-S1 purified by 2 M of urea and Vero E6 cell lysate. (A) GST-S1 withVero E6 cell lysate, **(B)** GST-S1 without lysate and **(C)** GST with lyate was used for pull-down assay. Red circles indicate GST-S1 and GST while green circles represent spots that are detectable in the gel using GST-S1 and Vero E6 lysate. H: High range marker.

6.3 Identification of Putative Interacting Partners by MALDI-TOF-TOF

The spots of putative interacting partners using Vero E6 cell lysate were isolated and digested by trypsin. After desalting by Ziptip, peptide masses were measured by MADTI-TOF-TOF. The PMFs were submitted to databases of NCBI, SWISS-Prot and TrEMBL for identification. Searching results showed that two of these spots were Peroxiredoxin 1 (Prx1) and Heat Shock 70 kDa Protein 5 (HSPA5) (Table 6.1). Their apparent masses on the 2D gels were same as their experimental masses, which were 26 kDa and 78 kDa respectively. On the other hand, several spots were identified as bacterial proteins including GroEL and Heat shock protein 40 (DnaJ). Also, some spots were identified as other mammalian and bacterial proteins spots but their M. W. and pI were not matched with the observed results. The identities of other spots could not be determined.

Table 6.1 – Summary of search results. Putative proteins with confidence interval (C. I.) of 90 or higher on figure 6.3 and 6.4 were shown. When the differences of M. W. and pI between putative proteins and apparent spots were less than 10 kDa and one respectively, the putative proteins were bolded. The proteins in red colour represent the positive controls.

Spot number	Putative protein	Accession number	Peptide count	Protein score	C. I.	Putative Protein		Apparent spot	
						M. W.	pI	M. W.	pI
Figure 6.2 A									
1	/	/	/	/	/	/	/	60	9
2	/	/	/	/	/	/	/	60	9
3	/	/	/	/	/	/	/	60	9
4	/	/	/	/	/	/	/	60	9
5	/	/	/	/	/	/	/	25	9
6	**Peroxiredoxin 1 [*H. sapiens*]**	**gi|55959887**	**6**	**80**	**96.34**	**18.96**	**6.41**	25	7
7	**DnaJ [*E. coli*]** **Chaperone Hsp40 [*E. coli*]** CENP-E [*H. sapiens*]	**gi|15829269** **gi|16128009** gi|29865	**12** **11** 15	**132** **118** 60.4	**100** **100** 92.05	**41.02** **41.07** 31.19	**7.97** **7.98** 5.46	45	8
8	/	/	/	/	/	/	/	45	8
9	**DnaJ [*E. coli*]** **Chaperone Hsp40 [*E. coli*]** **Chaperone Hsp40, co-chaperone with DnaK [*E. coli*]** **DnaJ-class molecular chaperone with C-terminal Zn finger domain [*E. coli*]**	**gi|15829269** **gi|16128009** **gi|16128009** **gi|75209619**	**9** **8** **8** **8**	**97.6** **83.4** **83.4** **81.3**	**100** **99.99** **98.33** **97.29**	**41.02** **41.07** **41.07** **41.05**	**7.97** **7.98** **7.98** **7.97**	45	7.5
10	/	/	/	/	/	/	/	75	5
11	/	/	/	/	/	/	/	75	5
12	/	/	/	/	/	/	/	65	5
13	/	/	/	/	/	/	/	65	5
14	/	/	/	/	/	/	/	65	5

15	**BiP protein [*H. sapiens*]** **heat shock 70kDa protein 5 [*H. sapiens*]** putative DNA binding protein [*E. coli*]	**gi|6470150** **gi|16507237** gi|15832767	**11** **11** 4	**108** **107** 56.7	**100** **100** 95.00	**70.89** **72.29** 7.978	**5.23** **5.07** 9.81	80	5.5
16	/	/	/	/	/	/	/	25	6
17	/	/	/	/	/	/	/	25	6
18	/	/	/	/	/	/	/	30	5.5
19	**GST class-pi [*E. coli*]**	**gi|2495108**	**7**	**81.1**	**97.16**	**23.42**	**5.93**	25	5.5
20	Chain N, Revised Atomic Structure Fitting Into A Groel(D398a)-Atp7 [*E. coli*] **GroEL [*E. coli*]** 60 kDa chaperonin [*E. coli*] GroEL [*E. coli*] DLP12 prophage; conserved protein [*E. coli*]	gi|90108999 **gi|18028158** gi|2493643 gi|38491462 gi|16128531	8 **7** 9 9 4	69.7 **63.3** 78.5 78.5 54.9	99.75 **98.91** 94.83163 94.83163 92.43	57.09 **52.00** 57594.97 57303.65 0	4.81 **4.79** 5.22 4.84 6.49	45	5
21	Chain N, Structural Basis For Groel-Assisted Protein Folding [*E. coli*] GroEL [*E. coli*] 60 kDa chaperonin [*E. coli*] unnamed protein product [*E. coli*] antigen of the monoclonal antibody Ki-67 [*H. sapiens*] cell proliferation antigen Ki-67, short form - human [*H. sapiens*] Antigen KI-67 [*H. sapiens*] GroEL [*E. coli*]	gi|29726368 gi|12519125 gi|6225121 gi|41614 gi|415821 gi|539556 gi|1170654 gi|38491472	11 11 12 6 20 19 19 10	91.7 91.6 107 77.3 77.2 68.9 61.3 77.9	100.00 100.00 99.99 99.96 99.83 98.88 93.54 94.07	57.03 57.28 56.62 9.14 319.30 319.25 358.53 57.29	4.84 4.81 4.85 9.4 9.51 9.49 9.45 4.79	40	5

22	GroEL [*E. coli*]	/	/	/	/	/	/	40	5
23	anti-HIV-1 gp120 antibody p20 heavy chain variable region [*H. sapiens*]	gi|1785872	5	60.8	92.75	14.10	9.2	35	5
24	/	/	/	/	/	/	/	35	4.5
25	/	/	/	/	/	/	/	35	4.5
26	/	/	/	/	/	/	/	25	5
27	S1 protein	gi|92429689	11	98.6	99.95	79.97	5.85	100	7
	spike glycoprotein	gi|42741270	8	76.7	92.18	50.30	6.32		
	spike glycoprotein	gi|42741279	8	76.6	92.00	50.46	6.32		

Figure 6.2 D									
1	**GroEL [*E. coli*]**	**gi|38491462**	**16**	**142**	**100**	**57.30**	**4.84**	60	5
3	/	/	/	/	/	/	/	55	5.5
4	Maturase homolog [*E. coli*]	gi|628756	6	69.5	99.74	14.55	10.79	55	5.5
	Curlin genes transcriptional activatory protein [*E. coli*]	gi|26106683	6	65.2	99.29	15.63	5.67		
	Brain specific protein [*H. sapiens*]	gi|5531807	6	60.7	92.58	18.55	9.54		
5	/	/	/	/	/	/	/	55	5.5
6	Maturase homolog [*E. coli*]	gi|628756	6	63.9	99.05	14.55	10.79	55	5
7	/	/	/	/	/	/	/	50	5
8	TBC1D7 protein [*H. sapiens*]	gi|13937893	7	67.3	98.38	33.95	8.07	50	5.5
	GroEL [*E. coli*]	**gi|18028158**	**7**	**58.6**	**96.77**	**52.00**	**4.79**		
	Curlin genes transcriptional activatory protein [*E. coli*]	gi|26106683	5	56.1	94.26	15.63	5.67		
9	Zinc finger protein 529 [*H. sapiens*]	Gi|24308255	7	59.9	91.08	53.66	9.05	45	5
10	/	/	/	/	/	/	/	45	5.5
11	/	/	/	/	/	/	/	40	5
12	FLJ00133 protein [*H. sapiens*]	gi|18676472	10	71	99.31	13.81	6.54	35	4.5
	SNED1 protein [*H. sapiens*]	gi|20379832	8	65.1	97.31	65.31	8.89		
	Hypothetical protein c3199 [*E. coli*]	gi|26109448	5	54.9	92.43	25.93	10.66		
13	Fused tRNA nucleotidyl transferase [*E. coli*]	gi|16130952	8	67.3	99.56	46.44	6.19	35	5
	Glucose-6-phosphate 1-dehydrogenase [*E. coli*]	gi|15831816	8	59.6	97.44	55.70	5.56		
	GroEL [*E. coli*]	gi|18028158	7	52.9	88.00	52.00	4.79		

14	TBC1 domain family, member 7 [*H. sapiens*] TBC1D7 protein [*H. sapiens*] Hypothetical protein b0387 [*E. coli*] a.k.a. mopA [*E. coli*]	gi|56205155 gi|13937893 gi|90111127 gi|146269	9 9 5 5	95.1 89.1 55.9 53.8	99.89 99.55 93.99 90.25	29.15 33.95 16.96 16.42	8.43 8.07 5.58 9.05	35	5
15	ORF; Unknown function	gi|12514511	5	55.5	93.41	12.00	10.28	35	5
16	GST	gi|71556047	8	85.5	98.97	23.52	5.6	35	5
17	/	/	/	/	/	/	/	30	5
18	/	/	/	/	/	/	/	30	5
19	Testis expressed sequence 12 [*H. sapiens*]	gi|13775182	5	59.7	90.67	14.10	5.22	30	5
20	/	/	/	/	/	/	/	35	5.5
21	ZNF610 protein [*H. sapiens*]	gi|76780061	8	66.7	98.14	53.41	9.33	35	5.5
22	S1 protein	gi|92429689	5	47.4	0	79.97	5.85	100	6
23	GST [*E. coli*]	gi|1699061	9	125	100.00	28.41	5.85	25	6.5
24	RP11-244N9.5 [*H. sapiens*] GST [*E. coli*]	gi|55958045 gi|2495108	6 7	61.3 74.4	93.54 86.72	18.99 23.42	8.17 5.93	25	6.5

Figure 6.3 A									
1	/	/	/	/	/	/	/	120	6.5
2	/	/	/	/	/	/	/	110	6.5
3	/	/	/	/	/	/	/	110	6.5
4	/	/	/	/	/	/	/	110	6.5
5	/	/	/	/	/	/	/	110	5.5
6	/	/	/	/	/	/	/	120	5.5
7	/	/	/	/	/	/	/	110	5.5
8	/	/	/	/	/	/	/	65	5
9	/	/	/	/	/	/	/	45	6
10	/	/	/	/	/	/	/	35	7
11	/	/	/	/	/	/	/	35	6
12	/	/	/	/	/	/	/	40	5
13	/	/	/	/	/	/	/	35	5.5
14	/	/	/	/	/	/	/	35	10
15	/	/	/	/	/	/	/	30	11
16	/	/	/	/	/	/	/	30	8
17	/	/	/	/	/	/	/	30	6
18	/	/	/	/	/	/	/	30	5.5
19	/	/	/	/	/	/	/	25	6.5
20	/	/	/	/	/	/	/	25	6.5
21	**GST class-pi**	**gi|2495108**	**5**	**76.9**	**92.53**	**23.42**	**5.93**	25	6.5
22	/	/	/	/	/	/	/	25	6.5
23	/	/	/	/	/	/	/	25	6.5
24	S1 protein S1 protein S1 protein	gi|92429689 gi|92698735 gi|89474571	11 12 11	104 91 83.7	99.99 99.71 98.44	79.97 139.11 139.03	5.85 5.52 5.56	100	6
25	GST class-pi	gi|2495108	7	110	100.00	23.42	5.93	25	6.5

Discussion

7.1 S Protein Expression in *E. coli*

7.1.1 Improving Recombinant Protein Expression Level and Solubility

Being an expression host, *E. coli* offers a means for the rapid and economical protein production with a high expression level, for instance, the soluble GST yield was more than 20 mg/L. Raising the bait (GST-S1) amount in pull-down assay is beneficial to prey interacting partners with low binding affinity since the interactions between the bait and the interacting partners are enhanced (Wong *et al.*, 2004). Besides pull-down assay, it can also be used for other studies that require huge protein amount, e.g. structure determination. Nonetheless, the yields of purified S protein fragments, especially S1 and S2, were extremely low (Figure 5.3). Several parameters including fusion tags, host strains, temperature, expression duration and lysis buffer contents potentially affecting solubility were hence optimized in this project.

GST- and His-MBP-tags were examined as they not only stabilize the recombinant protein but also increase the protein solubility (Sachdev *et al.*, 1999; Terpe, 2002). These tags can bind to glutathione and amylose respectively by their specific conformations and hence be immobilized and purified (Di Guan *et al.*, 1988; Smith *et al.*, 1988). His-tag was tried because this short tag with six histidine residues minimizes the chance of altering recombinant protein structure and function (Hakansson *et al.*, 2000; Terpe, 2002; Wu *et al.*, 1999). Another reason is His-fusion protein can be purified under denaturing conditions, that do not influence interaction between His-tag imidazole ring and transition metal ions on matrix (Hochuli *et al.*, 1987). Additionally, combination of His- and MBP-tag for purification and

stabilization respectively is widely used for the expression of many recombinant proteins (Hamilton *et al.*, 2002; Podmore *et al.*, 2002).

Level and solubility of protein expressed in various bacterial strains with their own characteristics and modifications can be altered (http://www.novagen.com/SharedImages/Novagen/05_PROEXP.pdf; http://www.embl.de/ExternalInfo/protein_unit/draft_frames/flowchart/exp_e_coli/Expression_Ecoli.html). First of all, BL21 lacking *lon* and *ompT* proteases is the most widely used expression host. For a protein encoded by bacterial rare codons, co-expression with tRNAs for AGG, AGA, CGG, AUA, CUA CCC and GGA in Rosetta 2 may reduce premature termination and thus increase the expression level. On the other hand, C41, a BL21 strain with unknown mutations, was effective to express certain toxic and membranous proteins (Chow *et al.*, 2005; Dumon-Seignovert *et al.*, 2004; Miroux *et al.*, 1996). Thus, it may help to express apoptosis-inducing S protein with a number of hydrophobic regions. Lastly, Origami with thioredoxin reductase and glutathione reductase mutants provides a more oxidizing cytoplasmic environment, which greatly facilitates disulfide bond formation and hence proper folding of S recombinant proteins.

Expression temperature and time are also important factors influencing expression level and solubility. Higher temperature and longer expression time may increase expression level as well as the chances of degradation and aggregation due to misfolding, and vice versa (Liao, 1991; Strandberg *et al.*, 1991). Their effects on S protein expression are unknown and therefore examined in this project.

Using an appropriate buffer can maintain the recombinant protein in the proper conformation and eliminate non-target contaminants. The buffer with high DTT concentration (100 mM) breaks disulphide bonds, preventing protein

misfolding due to abnormal disulphide bonding formation. Low concentrations of denaturing agents and detergents, such as 1 M of urea and 1 % (v/v) of triton X-100 respectively, help to extract target protein and remove non-specific binding proteins. At last, PBS closing to physiological conditions of human was examined because it is widely used in similar experiments.

Experimental results revealed solubility was increased when S protein fused with GST-tag, expressed at a lower temperature and lyzed in buffer with triton X-100, while the purity was improved after lysis in buffer with urea followed by affinity chromatography (Figure 5.3, 5.5 and 5.6). An unpredictable result was the protein lyzed in PBS had the highest solubility (Figure 5.6). A possible explanation is PBS with a lower salt concentration (137 mM) causes reduction of interactions between charged groups of protein molecules.

After optimization, only 100 μg/L of soluble GST-S1 was isolated (Figure 5.6). Fragment sizes, fragment locations and expression conditions are varied in other research groups but expression levels and solubility were similar. Most of the fragments were expressed in insoluble form (Chen *et al.*, 2005; Wei *et al.*, 2005; Yu *et al.*, 2003; Zhou *et al.*, 2004; Zuo *et al.*, 2005). Till now, a. a. residue 260 – 600, 397 – 796, RBD and HR, but not S1 and S2, were expressed in native conformations (Chen *et al.*, 2005; Hakansson-McReynolds *et al.*, 2006; Wei *et al.*, 2005; Xu *et al.*, 2004; Xu *et al.*, 2004). By refolding of inclusion bodies, soluble full-length, S2, HR and a. a. residue 450 – 650 were produced but native folding was not yet confirmed (Ho *et al.*, 2004; Hsu *et al.*, 2004; Zhao *et al.*, 2005; Zuo *et al.*, 2005).

7.1.2 S Recombinant Protein Bound by GroEL

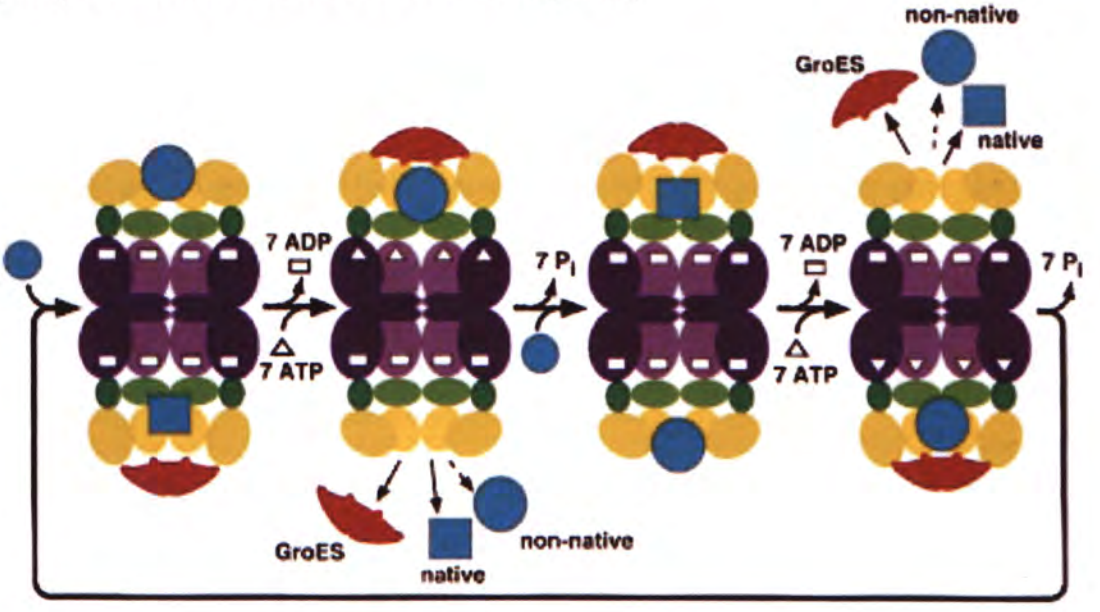
All expressed GST-S protein fragments in various conditions were bound by GroEL, which was irremovable by affinity chromatography (Figure 5.3 and 5.5). Similar problems were also encountered by another research group (Yu *et al.*, 2003). In fact, GroEL is an endogenous bacterial chaperone preventing irreversible protein aggregation as well as assisting protein folding into the native conformation (Figure 7.1) (Sakikawa *et al.*, 1999; Taguchi *et al.*, 2005, Viitanen *et al.*, 1992). It exists in the form of a double ring of seven subunits each with a central cavity (Braig *et al.*, 1994) and then links with a cofactor GroES to construct a GroEL/ES complex (Saibil *et al.*, 2002; Sigler *et al.*, 1998). Hydrophobic residues of entrance of the central cavity interact with the hydrophobic residues of non-native substrate exposed on the surface, followed by ATP hydrolysis promoting folding, and finally releasing the substrate (Horwich *et al.*, 1993; Taguchi *et al.*, 2005; Van Dyk *et al.*, 1989).

Assistance of protein folding by GroEL is a common phenomenon because approximately 10 % of newly synthesized polypeptides in *E. coli* were the substrates (Ewalt *et al.*, 1997). A characteristic of these substrates is preferentially presences of αβ-fold, which is composed of an α-helix and buried β-sheet (Houry *et al.*, 1999). The secondary structure prediction results revealed that all S protein fragments contain α-helix and β-sheet except RBD (a. a. 318 – 510), which is confirmed by crystal structure analysis of RBD (Li *et al.*, 2005). Hence, the interaction between RBD and GroEL is possible to be induced by exposure of other hydrophobic motif or patch (Chaudhuri *et al.*, 2005; Stan *et al.*, 2006).

Generally, the substrates were simply released in the presence of ATP but release of S protein by ATP failed in our project (Figure 5.10) (Grantcharova *et al.*, 2001; Thirumalai *et al.*, 2001). Besides misfolding of S proteins, the failure suggests

that S proteins binding to GroEL is folded properly and thus functional (Venkatesh *et al.*, 2004). Thus, pull-down assays using both GST-S1 with GroEL and GST-S1 purified by 2 M of urea were performed.

Figure 7.1 – Schematic diagram of folding mechanism by GroEL/ES complex (Adapted and modified from Feltham *et al.*, 2000).



7.2 S Protein Expression in *P. pastoris*

7.2.1 Advantages of Using P. pastoris

P. pastoris is another high-yield, rapid and economical system for a wide variety of recombinant protein production. As a eukaryote, *P. pastoris* can also post-translational modified the expressed protein such as folding, disulfide bond formation, glycosylation and proteolytic processing. The protein structure and modifications are similar to that expressed in mammalian cells, for example, the core structure of the protein N-linked glycan formed in *P. pastoris* is same as that of mammalian cells (Tarentino *et al.*, 1985). Therefore, the proteins expressed in *P. pastoris* should have a higher chance to be in native conformation and functional when compared with proteins expressed in the bacterial system.

Secretion of recombinant protein is another great advantage of *P. pastoris*. Except low levels of endogenous proteins, the heterologous proteins with a signal peptide are only secreted to the medium and comprise the vast majority of the total protein. This would minimize purification steps and improve recovery yield. Moreover, low levels of endogenous proteins including proteases reduce the chance of protein degradation as well.

Various yeast hosts are suitable for foreign protein expression but *P. pastoris* possesses its own benefits. Firstly, alcohol oxidase I (*AOXI*) gene is methanol-inducible and tightly regulated. Before induction, energy is therefore saved for cell growth rather than lost for leakage of protein expression. Moreover, *P. pastoris* can be cultivated in a high density (100 g/L) and MeOH is utilized for the expression induction, so that space and cost required are relatively low.

Hyperglycosylation is a common problem in various yeast hosts. *P. pastoris* is more beneficial than other hosts such as *S. cerevisiae*, since the longest N-linked

glycans in *P. pastoris* is approximately 8-14 mannose residues while that in *S. cerevisiae* is between 50 and 100 mannose residues (Cregg *et al.*, 1993; Daly *et al.*, 2005). The glycoprotein structure and function expressed in *P. pastoris* is more similar and compatible to that expressed in mammalian systems.

7.2.2 Variation of S Fragment Expression Levels

Only αA-(318-510) level reached 46mg/L while the levels of other recombinant proteins expressed in the same conditions were much lower. Typically, only one-tenth of the amount of αA-(318-510) or less could be obtained (Figure 5.18 and 5.19). A possible reason leading to the difference of expression levels of four S protein fragments in *P. pastoris*, but not in *E. coli*, is the codon bias. Expression level of a protein encoded by rare codons can be increased by five to ten folds after codon optimization (Hu *et al.*, 2006; Xiong *et al.*, 2005; Xiong *et al.*, 2006).

Nucleotide sequence encoding S protein was submitted to two databases for rare codon analysis. Even two databases showed little diverse codon frequencies, both results revealed the expression levels are negatively related to the number of rare codons (Table 7.1). αA-(318-510) contains the lowest rare codon number, followed by αA-(903-1187) and αA-(15-317). αA-(587-826) is the only exceptional case, in which the number of rare codons was similar to that of αA-(15-317) but the expression level was much lower (Figure 5.19).

When only the lowest frequencies (2.0 or 2.2 / 1,000) in the database "Rare codons' Search" are considered, αA-(587-826) contains two of these codons while other fragments contain one or less. The numbers of these rare codons are negatively related to their expression levels as well. Collectively, the analysis suggested the expression levels of S protein fragments could be improved after codon optimization.

Table 7.1 – Rare codons of S protein expressed in *P. pastoris*. The nucleotide sequencing of S protein was analyzed by "Rare codons' Search" (http://molbiol.ru/eng/scripts/01_11.html) and "Graphical Codon Usage Analyzer" (http://gcua.schoedl.de/), that are represented by database 1 and 2 respectively. The total frequency is 1,000. Bold numbers in database 1 indicate the frequencies are 2 or 2.2.

no. of a. a. residues	Codon	a. a. residues encoded	Frequency (Database 1)	Frequency (Database 2)	a. a. residues of S protein	
18	cgg	Arg	**2**	5	S1	15-317
19	ugc	Cys	4.2	/		
39	ggg	Gly	/	10		
68	ggg	Gly	/	10		
113	ucg	Ser	/	8		
126	cga	Arg	4.4	10		
159	ugc	Cys	4.2	/		
169	ucg	Ser	/	8		
183	cga	Arg	4.4	10		
192	ggg	Gly	/	10		
194	cuc	Leu	/	8		
264	cuc	Leu	/	8		
286	cuc	Leu	/	8		
288	ugc	Cys	4.2	/		
292	agc	Ser	/	9		
355	cuc	Leu	/	8		318-510
366	ugc	Cys	4.2	/		
378	ugc	Cys	4.2	/		
398	gcg	Ala	3.8	5		
467	ugc	Cys	4.2	/		
507	ccg	Pro	4.2	10		
532	cuc	Leu	/	8		
563	cga	Arg	/	10		
563	cga	Arg	4.4	/		
576	ugc	Cys	4.2	10		
603	ugc	Cys	4.2	/		587-826
615	cuc	Leu	/	8		
620	cgc	Arg	**2.2**	6		
648	ugc	Cys	4.2	/		
670	agc	Ser	/	9		
703	agc	Ser	/	9	S2	
725	ugc	Cys	4.2	/		
736	cuc	Leu	/	8		
740	agc	Ser	/	9		
742	ugc	Cys	4.2	/		
749	cuc	Leu	/	8		
758	cgc	Arg	**2.2**	6		
804	cuc	Leu	/	8		
810	cuc	Leu	/	8		
822	ugc	Cys	4.2	/		
831	cuc	Leu	/	8		
834	gcg	Ala	3.8	5		
847	cuc	Leu	/	8		
898	cuc	Leu	/	8		
902	caa	Gln	/	5		
912	gcg	Ala	3.8	/		903-1187
949	agc	Ser	/	9		
964	ucg	Ser	/	8		
965	cga	Arg	4.4	10		
971	gcg	Ala	3.8	5		
985	agc	Ser	/	9		
1039	ccg	Pro	4.2	10		
1060	gcg	Ala	3.8	5		
1167	cgc	Arg	**2.2**	6		
1168	cuc	Leu	/	8		
1179	cuc	Leu	/	8		

GC content is another common factor leading to variation of expression levels in *P. pastoris* (Kotula *et al.*, 1991; Sreekrishna *et al.*, 1997). Genes with high AT content may be ineffectively transcribed because of premature termination (Romanos *et al.*, 1992). Nevertheless, ratios of GC/AT of all S protein fragments were comparable; indicating that the expression level variation is contributed by other factors rather than GC content (Table 7.2).

Table 7.2 – GC contents of S protein fragments. No significant difference was detected between the GC contents of these fragments.

a. a. residue of S fragments	A (%)	T (%)	AT (%)	G (%)	C (%)	GC (%)	GC/AT ratio
13-672 (S1)	27.3	35.3	62.6	18.2	19.2	37.4	0.597
680-1192 (S2)	30.1	29.7	59.8	18.8	21.4	40.2	0.672
15-317	27.4	36.2	63.6	17.6	18.8	36.4	0.572
318-510	26.9	35.4	62.3	18.7	19	37.7	0.605
587-826	28.5	32.5	61	18.3	20.7	39	0.639
903-1187	31.8	27.8	59.6	18.5	21.9	40.4	0.678

7.2.3 Sizes of S Protein Fragments

Results of SEC purification of αA-(318-510) showed that the protein native size was between 27 and 478 kDa, while the peak height was in the location of 60 kDa (Figure 5.24). Its apparent molecular weight determined by SDS-PAGE was between 34 and 100 kDa (Figure 5.21 A). By comparison, αA-(318-510) was a monomeric protein but the apparent size of the glycosylated protein determined by SEC was higher than that determined by SDS-PAGE (Figure 5.24 and 5.25). Putative explanations include variations of glycan lengths, as well as the glycans in extended form rather than globular form. The native size was further confirmed after deglycosylation, however the protein became aggregated (Figure 5.26). This might be caused by the removal of carbohydrates, which worsens protein stability and leads to protein aggregation (Endo *et al.*, 1992; Jafari-Aghdam *et al.*, 2004; Wang *et al.*, 1996).

Sizes of de-N-glycosylated αA-(15-317), αA-(318-510) and αA-(903-1187) were still 5 – 10 kDa higher than expected and a few extra bands or smears were detected by SDS-PAGE (Figure 5.21). A probable reason is the presence of several O-linked mannose residues linking to serine or threonine residues. Another reason is the imconplete de-N-glycosylation by PNGase F. A sharper band may be detected if N-linked glycans are removed by both PNGase F and other N-glycanases such as Endo H and Endo F (Molhoj *et al.*, 2001).

7.3 Identification of Interacting Partners

7.3.1 Relationship between S Protein and Putative Interacting Partners

Two mammalian proteins, PrxI and HSPA5, were preyed by using GST-S1 as bait (Figure 6.3). PrxI, a member of Prxs, is a thiol-specific antioxidant protein of 26 kDa located at cytosol and nucleus. By oxidizing two active sites of cysteines to sulfenic acids, hydroperoxide groups (ROOH) of the substrate are reduced to alcohol groups (ROH). GST was proved to interact with another member containing only one active site (Manevich *et al.*, 2004; Ralat *et al.*, 2006). Furthermore, PrxI is present in the cytosol and the nucleus but not extracellular membrane. Therefore, PrxI is predicted to bind to GST-tag instead of being a receptor or interacting partner of S protein.

Another preyed mammalian protein is HSPA5, or called glucose-regulated protein 78 (Grp78), which is a member of Hsp70 family located in ER (Haas, 1994; Wei *et al.*, 1996). In cooperation with other chaperones including Grp94 and Calnexin, the major function of HSPA5 is folding, assembly and secretion of protein by association with shot, extended and exposed hydrophobic residues followed by ATP hydrolysis (Helenius, 1994; Wei *et al.*, 1996). The major substrate is heavy chains of antibodies as well as other proteins containing short hydrophobic residues (Blond-Elguindi *et al.*, 1993; Flynn *et al.*, 1991; Hass *et al.*, 1983). HSPA5 is predicted to be an interacting partner of S rather than a functional receptor as HSPA5 is located in ER and is a chaperone likes GroEL. Since S was bound by GroEL during expression in *E. coli*, the interaction between HSPA5 and S may be due to exposed hydrophobic residues of the S protein.

Besides the mammalian proteins, a few spots of 30 – 40 kDa were identified as GroEL and its subunits, as well as DnaJ, another bacterial chaperone (Figure 6.3).

Proteases presented in the lysate may be a possible answer explaining why GroEL with smaller sizes were present in 2D gel using GST-S1 and Vero E6 lysate (Figure 6.3). The proteases digest GroEL during pull-down assay but the cleaved GroEL products still bind to GST-S1 and were therefore detected in the 2D gel.

PrxI and HSP5A were only present in one of the trials of pull-down assay, and the patterns of spots in various trials were not identical (Figure 6.3 and 6.4). Slight variation of the confluence and passage between batches of Vero E6 cell lysate may be one of the explanations for this inconsistency. Experimental errors might explain the difference between trials as well.

7.3.2 Failure of Finding ACE2

ACE2, a 120 kDa protein, was the functional receptor of SARS-CoV and should be found in the 2D gels generated. However, only three chaperones were identified while no spots could be observed in the expected location for ACE2, suggesting that ACE2 could not be detected by using GST-S1 with 2D gel electrophoresis (Figure 6.3 and 6.4). However, ACE2 could be detected by western blotting using αA-RBD as bait, suggesting that the condition of pull-down assay is capable to capture interacting partners (Figure 6.2).

Aggregation of GST-S1 suggested that the protein might be misfolded, leading to the loss of function. The interaction between GST-S1 and GroEL could also cause the hindrance of the binding site because GroEL binds to GST-S1 was in the form of cage-like structure. On the other hand, absence of ACE2 could be caused by the insensitivity of silver staining. The detection limit of silver staining is in nanogram level while that of western blotting is in pictogram level. It was possible that ACE2 was already captured by GST-S1 but the low protein level led to

the failure of detection by silver staining. Taken together, misfolding of GST-S1 and insensitivity of silver staining might be the most probable reasons causing the failure.

7.3.3 Difficulty in the Identification of Protein Spots

Searching results of other spots only showed a number of proteins with low C. I. and protein score. Coverage and the number of peaks matching to these proteins were low as well, thus most of samples were unidentified (Table 7.2). Major reasons causing such failure is related to the sample preparation and the errors from the MALDI-TOF-TOF technique.

Protein amount is an important criterion determining whether identification is successful or not. In general, the identification of a spot or band with a larger amount is easier because a higher amount of protein leads to a higher peak intensity, resolution and signal-to-noise (S/N) ratio (Table 7.3). In our experience, some peptides can only be detected when protein levels are high. Moreover, more matching peptides increase the chance of spot identification. Sufficient amounts of proteins improve the numbers and intensities of peptides subjecting to MS/MS, leading to more accurate results. Hence, the recombinant proteins (GST-S1 and GST) and bacterial proteins (GroEL), but not other proteins with a low level from Vero E6 cells, could be identified in most of our 2D gels. One of the methods that can solve this problem is to pool protein spots from many trials of experiments in order to increase the protein amounts.

On the other hand, more than 50 spots were presented in a 2D gel using GST-S1 as bait (Figure 6.3 and 6.4). There will be a chance that a protein spot contains a few proteins rather than a single protein. PMF containing more than

one set of peptides is more difficult to be analyzed. Using two or more 2D gels with various percentages, such as 8 % and 15 % might resolve the spots.

A possible reason, which can explain why only recombinant proteins and bacterial proteins but not proteins from Vero E6 cells could be identified, is the post-translational modifications of Vero E6 cell proteins. Any modification changing peptide masses may lead to the failure in spot identification. Even if partial modifications are considered before searching, some complicated modifications such as glycosylation are difficult to be predicted. Furthermore, probability of glycosylated peptide ionization is relatively lower and thus the number of peptides generated will be reduced. Moreover, the inclusion of unrelated modifications during the searching process may have a negative effect on the searching results.

Table 7.3 – Protein scores and numbers of matched peaks of selected proteins. (A) Both the score and the number of matched peaks of putative protein were high hence the protein was identified. **(B)** Low score and number of matched peaks of the putative protein caused identification failure.

(A)

	Accession Number	Protein Name	Protein PI	Protein	Peptide Count	Protein Score	Protein Score C.I.	Total Ion Score
1	gi|92429689	S1 protein [SARS coronavirus GD322]	5.8500	79974.2	11	99	99.949	
2	gi|42741270	spike glycoprotein [SARS coronavirus TW-GD4]	6.3200	50299.7	8	77	92.177	
3	gi|42741279	spike glycoprotein [SARS coronavirus TW-YM2]	6.3200	50456.8	8	77	91.995	
4	gi|89474492	spike glycoprotein [SARS coronavirus]	5.6200	139034.9	11	74	85.765	
5	gi|32454346	spike glycoprotein [SARS coronavirus Shanhgai LY]	5.5100	138762.7	11	74	85.094	
6	gi|88860264	lysyl-tRNA synthetase [Pseudoalteromonas tunicata D2]	5.0600	58818.5	8	72	76.914	
7	gi|71144218	conserved hypothetical protein [Colwellia psychrerythraea 34H]	9.8900	13716.6	5	65	0.000	
8	gi|5230642	movement protein [African cassava mosaic virus-[Cameroon]]	6.6700	33619.0	6	63	0.000	
9	gi|99078937	S protein [SARS coronavirus]	5.6000	139009.8	10	63	0.000	
10	gi|68226738	hypothetical protein LOC212281 [Mus musculus]	9.4000	111407.0	9	63	0.000	
11	gi|44894077	spike glycoprotein [SARS coronavirus cw037]	8.0300	30600.2	6	63	0.000	
12	gi|72145970	PREDICTED: similar to CENP-F kinetochore protein (Centromere protein F) (Mitosin) [AH	4.8500	415820.2	15	62	0.000	
13	gi|44894075	spike glycoprotein [SARS coronavirus cw049]	8.3500	31573.7	6	62	0.000	
14	gi|50935555	hypothetical protein [Oryza sativa (japonica cultivar-group)]	9.3300	15860.0	5	62	0.000	
15	gi|10039039	aspartate carbamoyltransferase catalytic chain [Buchnera aphidicola str. APS (Acyrthosi	9.4900	35530.8	6	62	0.000	
16	gi|56554865	spike glycoprotein [SARS coronavirus A013]	5.5700	139111.8	10	62	0.000	
17	gi|4322668	retinoic acid receptor RXR [Schistosoma mansoni]	6.2100	82509.1	1			12
18	gi|6467333	elongation factor 1-a [Calliobothrium sp.]	7.2300	30103.5	1			8
19	gi|91082427	PREDICTED: similar to Centromeric protein E (CENP-E protein) [Tribolium castaneum]	5.4500	104993.0	1			8
20	gi|46123049	hypothetical protein FG05902.1 [Gibberella zeae PH-1]	5.7800	111909.7	1			8

(B)

	Accession Number	Protein Name	Protein PI	Protein	Peptide Count	Protein Score	Protein Score C.I.	Total Ion Score
1	gi|2465205	centromere protein A [Mus musculus]	12.6100	2325.3	2	41	0.000	
2	gi|90298108	hypothetical protein CIMG_08943 [Coccidioides immitis RS]	9.1500	7898.2	2	37	0.000	
3	gi|78496238	NAD binding site:amine oxidase [Rhodopseudomonas palustris BisB18]	10.4900	8261.5	2	35	0.000	
4	gi|69936516	hypothetical protein PdenDRAFT_1581 [Paracoccus denitrificans PD1222]	9.6200	6549.6	2	35	0.000	
5	gi|82746001	similar to cellobiose phosphotransferase enzyme IIB component [Clostridium beijerincki N	7.7100	11280.0	2	34	0.000	
6	gi|49328557	transposase for IS660, C-terminal region [Bacillus thuringiensis serovar konkukian str. 97	9.9200	14529.7	2	33	0.000	
7	gi|18874315	truncated envelope glycoprotein [Human immunodeficiency virus 1]	9.2400	9029.8	2	33	0.000	
8	gi|50878060	probable transposase (partial length) [Desulfotalea psychrophila LSv54]	9.5600	9755.0	2	33	0.000	
9	gi|51459982	iron-sulfur cofactor synthesis protein IscU/NifU [Rickettsia typhi str. Wilmington]	6.7300	14126.3	2	33	0.000	
10	gi|18874301	truncated envelope glycoprotein [Human immunodeficiency virus 1]	9.8800	9414.0	2	33	0.000	
11	gi|40068860	NEQ346 [Nanoarchaeum equitans Kin4-M]	9.7000	9356.2	2	32	0.000	
12	gi|71146177	hypothetical protein CPS_3366 [Colwellia psychrerythraea 34H]	9.1900	15558.1	2	32	0.000	
13	gi|21264187	hypothetical protein XACa0002 [Xanthomonas axonopodis pv. citri str. 306]	9.9100	9364.9	2	32	0.000	
14	gi|56416819	iron-sulfur cofactor synthesis protein [Anaplasma marginale str. St. Maries]	8.1900	14041.2	2	32	0.000	
15	gi|68127468	hypothetical protein, conserved [Leishmania major]	9.7100	10935.5	2	32	0.000	
16	gi|85716131	ABC transporter ATP-binding protein [Nitrobacter sp. Nb-311A]	7.9300	27625.7	2	31	0.000	5
17	gi|32400834	abscisic acid-induced protein [Triticum aestivum]	11.7400	10950.2	2	31	0.000	
18	gi|27381175	ABC transporter ATP-binding protein [Bradyrhizobium japonicum USDA 110]	6.6000	28363.9	2	31	0.000	5
19	gi|72396734	conserved hypothetical protein [Methanosarcina barkeri str. fusaro]	10.1400	12295.8	2	31	0.000	
20	gi|3668078	hypothetical protein [Arabidopsis thaliana]	9.1400	11271.0	2	31	0.000	

Table 7.4 – Cluster areas, peak areas and S/N ratios of selected proteins. The average cluster area, peak area and S/N ratio of matched masses of GST with larger protein amount were higher than that of HSPA5 with smaller protein amount.

GST (Spot #23 in Figure 6.3 A)				HSP5A (Spot #2 in Figure 6.3 A)			
Observed Mass	Cluster Area	Peak Area	S/N ratio	Observed Mass	Cluster Area	Peak Area	S/N ratio
2045.026	483700	160314	1584	1566.7715	51142	21783	314.1
2269.12	80758	23591	251.5	1815.9972	41154	15957	274.4
1516.811	65125	27039	366.3	2165.0042	20467	5705	148
1032.609	56507	30500	487.7	2235.1738	16137	4750	110.7
1801.945	54815	20696	228.5	1833.9081	15901	5666	88.1
2357.182	17759	5898	52.8	1512.7531	10387	4121	79.1
2326.023	10961	4265	34.2	986.5270	7020	7020	31.9
1182.615	5920	2805	47.8	1887.9775	6462	2305	39.5
1182.738	5850	3058	42.2	1934.0231	5315	2376	45.5
1094.585	5512	2982	45.3	1588.8052	4060	4060	24
				1645.8636	3188	3188	30.9

7.4 Conclusion

Spike protein S1 and S2 domains as well as four fragments were expressed in *E. coli* and *P. pastoris*. After expression condition optimization, GST-S1 was expressed in *E. coli* strain BL21, 16 ℃ overnight and then lyzed in PBS. The yield of soluble GST-S1 purified by affinity chromatography and SEC was approximately 100 μg/L. S protein associated with GroEL and some non-target proteins forms protein aggregates. GroEL, but not other unrelated proteins, was removed by the addition of 2 M urea in PBS. On the other hand, soluble αA-(318-510) with a yield of 46 mg/L was secreted from *P. pastoris* and the yield reached a maximum after 144 hours of induction. The production yields of other S domains and fragments were much lower. Native αA-(318-510) was a monomer but became aggregates after de-N-glycosylation. The result implied ACE2 could be captured by pull-down assay using native αA-(318-510).

Three interacting partners, PrxI, HSPA5 and DnaJ, from Vero E6 cell lysates were identified after pull-down assay using GST-S1 as bait, followed by 2D gel analysis and mass spectrometry; however, they may only interact with the GST-tag and S protein hydrophobic residues respectively instead of being functional receptors. Moreover, bacterial GroEL subunits were identified. They may be the digestion products of GroEL by protease in Vero E6 cell lysates. Lastly, some spots observed when GST-S1 as bait is putative interacting partners but their identities remain to be determined.

7.5 Future Prospects

GST-S proteins produced in *E. coli* were bound by GroEL. After optimizations and purifications, they were still associated with non-target proteins in form of aggregation. Moreover, effects of glycosylation and deglycosylation on αA-(15-317), αA-(318-510) and αA-(903-1187) were not yet understood. It is unknown whether these S proteins are in native conformation and functional. The pull-down assay will be performed in order to determine whether the S protein specifically interacts with ACE2 or other proteins. On the other hand, since αA-S1 and αA-S2 failed to be produced and insoluble respectively, the expression conditions such as lowering expression temperature, altering MeOH concentration and using strains other than KM71H will be optimized to increase both expression levels and solubility.

In order to identify the putative interacting partners successfully, several methods will be tried including the increase in the protein amount by repeating the pull-down assay, followed by pooling spots from 2D gels for mass spectrometric analysis. Also, 2D gels with various percentages will be performed to ensure that each spot contains only one protein instead a mixture of many proteins.

As S recombinant proteins were expressed in bacteria and yeast rather than the mammalian system, the post-translational modification and protein folding of different expression systems are varied. Consequently, if some putative functional receptors will be identified by pull-down assay, the interaction will be further validated by co-immunoprecipitation using S and the putative receptors expressed in the mammalian system.

Appendix

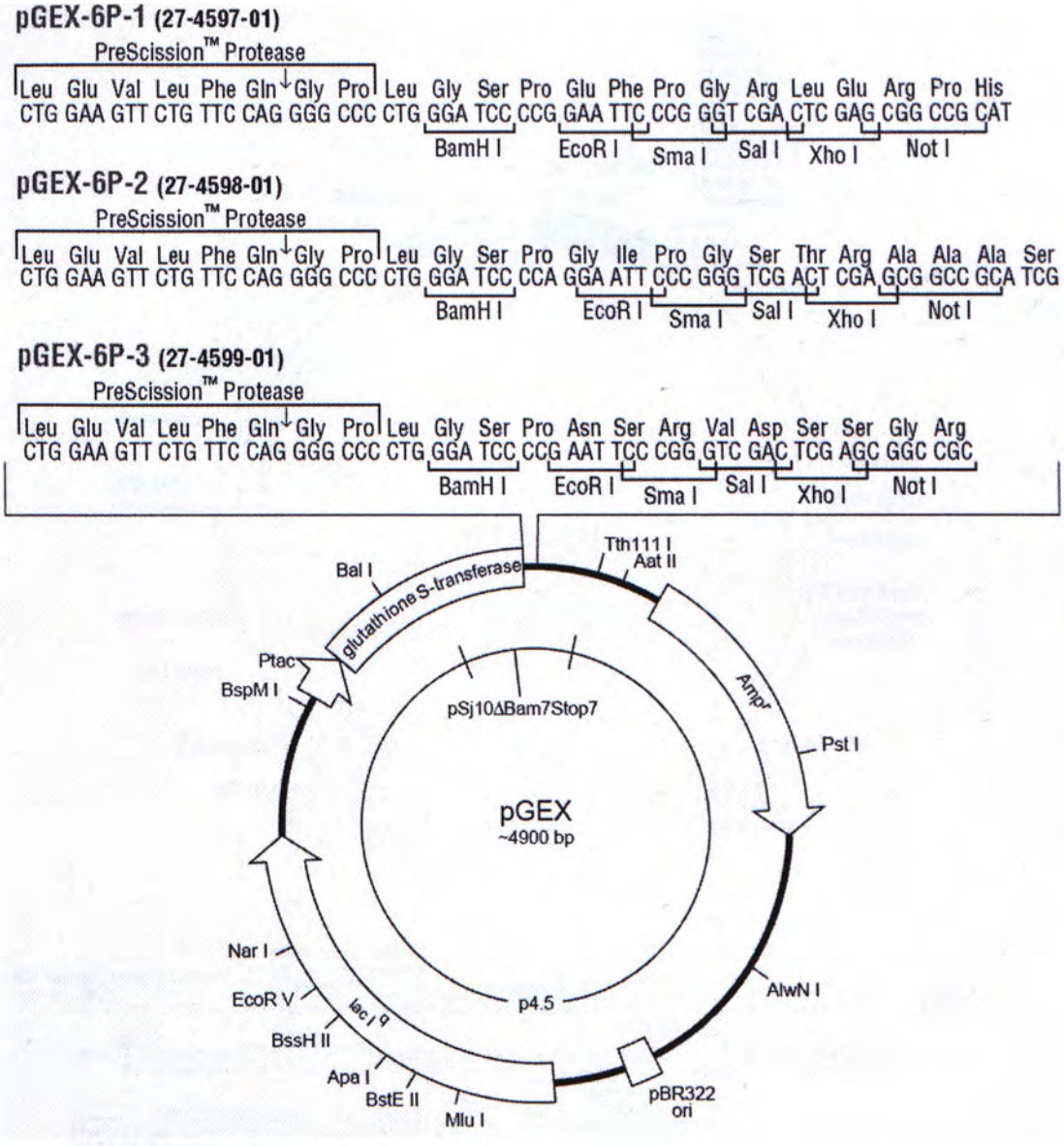


Figure 8.1 – Vector map of pGEX-6P-1 (Adapted and modified from http://www5.amershambiosciences.com/aptrix/upp01077.nsf/Content/Products?OpenDocument&parentid=27459701&moduleid=38859).

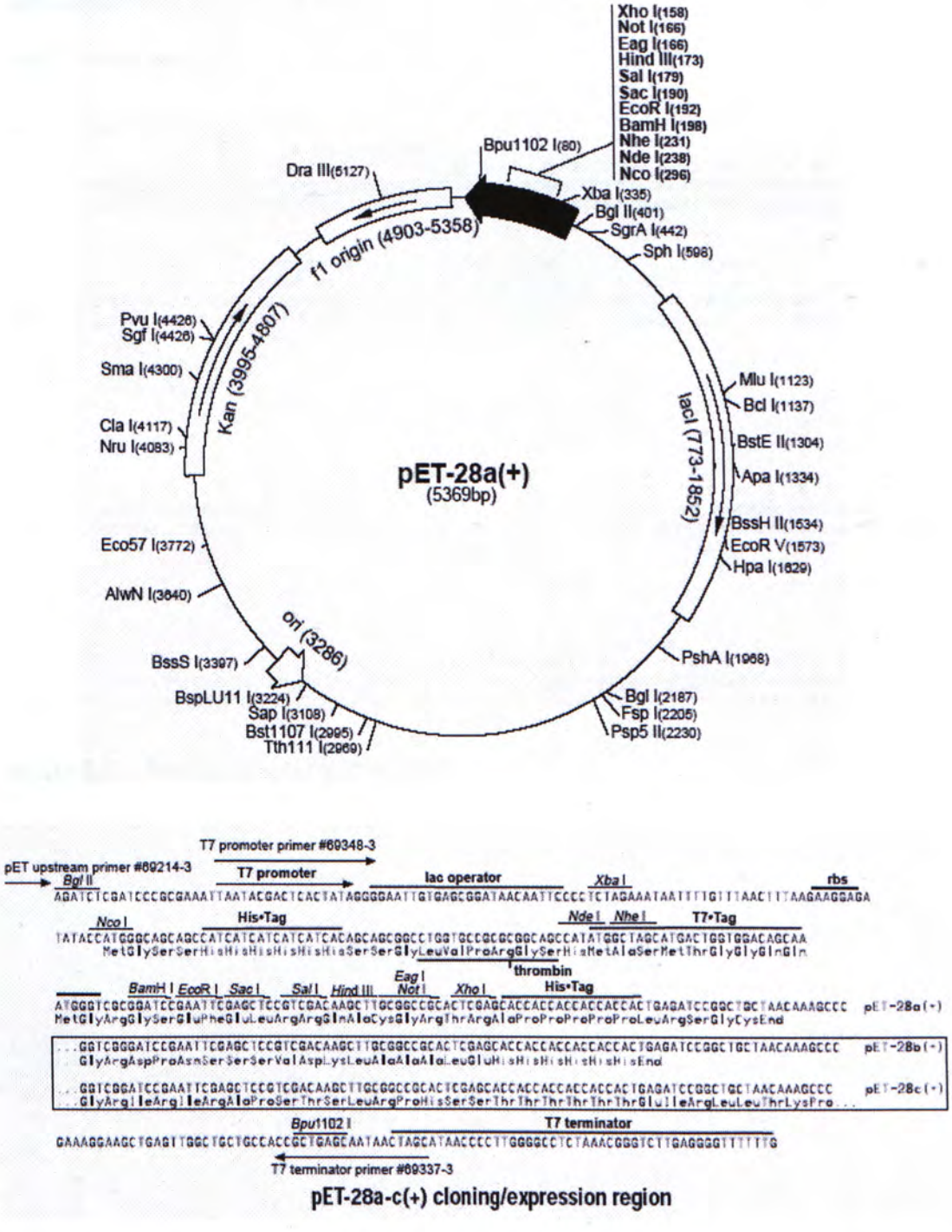


Figure 8.2 – Vector map of pET-28a (Adapted and modified from http://www.merckbiosciences.com/docs/NDIS/TB074-000.pdf).

pRHisMBP Restriction Map

4060 base pairs

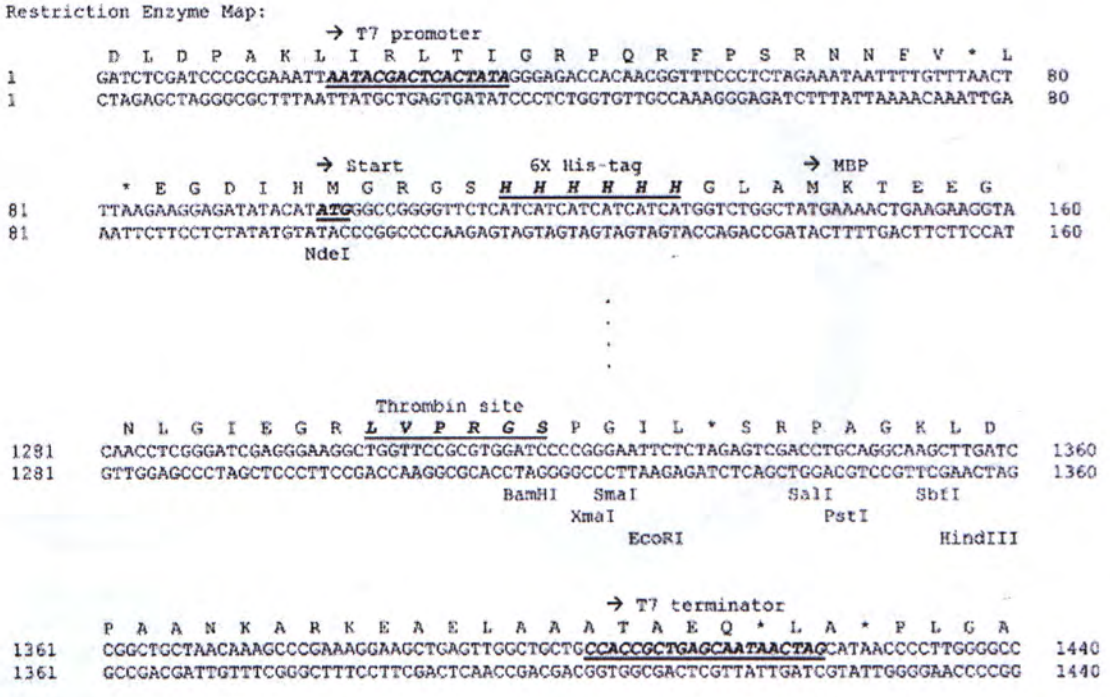
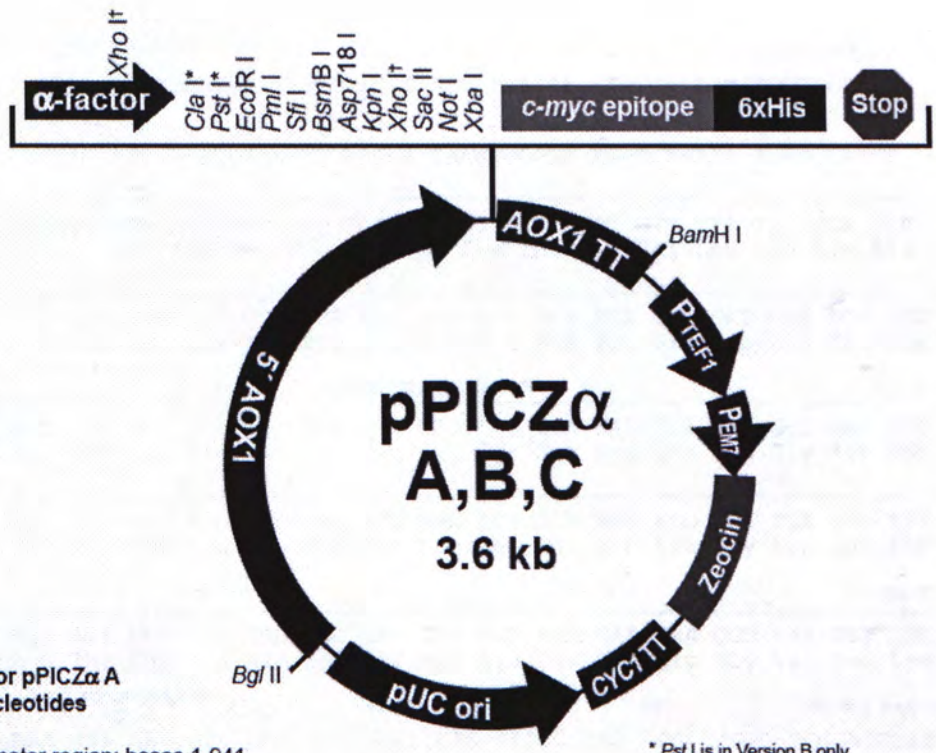


Figure 8.3 – Vector map of pRHisMBP.



Comments for pPICZα A
3593 nucleotides

5′ *AOX1* promoter region: bases 1-941
5′ *AOX1* priming site: bases 855-875
α-factor signal sequence: bases 941-1207
Multiple cloning site: bases 1208-1276
c-myc epitope: bases 1275-1304
Polyhistidine (6xHis) tag: bases 1320-1337
3′ *AOX1* priming site: bases 1423-1443
AOX1 transcription termination region: bases 1341-1682
TEF1 promoter: bases 1683-2093
EM7 promoter: bases 2095-2162
Sh ble ORF: bases 2163-2537
CYC1 transcription termination region: bases 2538-2855
pUC origin: bases 2866-3539 (complementary strand)

* *Pst* I is in Version B only
Cla I is in Version C only

†The two *Xho* I sites in the vector allow the user to clone their gene in frame with the Kex2 cleavage site, resulting in expression of their native gene without additional amino acids at the N-terminus.

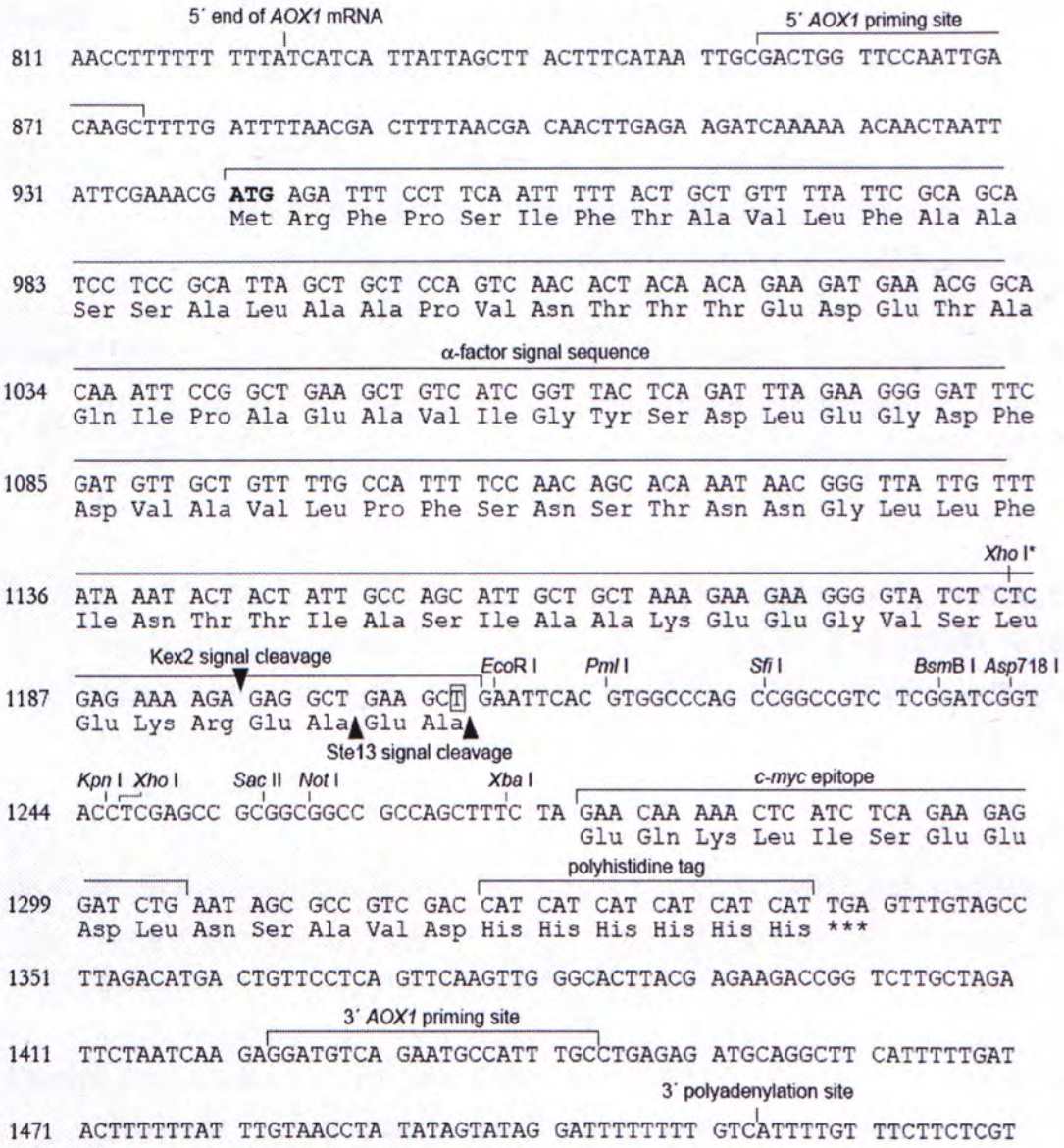


Figure 8.4 – Vector map of pPICZα-A (Adapted and modified from http://www.invitrogen.com/content/sfs/vectors/ppiczalpha_map.pdf and http://www.invitrogen.com/content/sfs/vectors/ppiczalphaa_mcs.pdf).

References

Averill, D. B., Ishiyama, Y., Chappell, M. C. & Ferrario, C. M. (2003). Cardiac angiotensin-(1-7) in ischemic cardiomyopathy. *Circulation* **108,** 2141-6.

Babcock, G. J., Esshaki, D. J., Thomas, W. D. Jr. & Ambrosino, D. M. (2004). Amino acids 270 to 510 of the severe acute respiratory syndrome coronavirus spike protein are required for interaction with receptor. *J Virol* **78,** 4552-60.

Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F. & Gething, M. J. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell* **75,** 717-28.

Bosch, B. J, Martina, B. E., Van Der Zee, R., Lepault, J., Haijema, B. J., Versluis, C., Heck, A. J., De Groot, R., Osterhaus, A. D. & Rottier, P. J. (2004). Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proc Natl Acad Sci U S A* **101,** 8455-60.

Broer, R., Boson, B., Spaan, W., Cosset, F. L. & Corver, J. (2006). Important role for the transmembrane domain of severe acute respiratory syndrome coronavirus spike protein during entry. *J Virol* **80,** 1302-10.

Centers for Disease Control and Prevention. (2003). Severe Acute Respiratory Syndrome --- Singapore, 2003. *Morbidity and Mortality Weekly report* **52,** 405-11.

Chan, P. K., To, K. F., Lo, A. W., Cheung, J. L., Chu, I., Au, F. W., Tong, J. H., Tam, J. S., Sung, J. J. & Ng, H. K. (2004a). Persistent infection of SARS coronavirus in colonic cells in vitro. *J Med Virol* **74,** 1-7.

Chan, P. K., Tam, J. S., Lam, C. W., Chan, E., Wu, A., Li, C. K., Buckley, T. A., Ng, K. C., Joynt, G. M., Cheng, F. W., To, K. F., Lee, N., Hui, D. S., Cheung, J. L., Chu, I., Liu, E., Chung, S. S. & Sung, J. J. (2003). Human metapneumovirus detection in patients with severe acute respiratory syndrome. *Emerg Infect Dis* **9,** 1058-63.

Chan, W. M., Kwan, Y. W., Wan, H. S., Leung, C. W. & Chiu, M. C. (2004b). Epidemiologic linkage and public health implication of a cluster of severe acute respiratory syndrome in an extended family. *Pediatr Infect Dis J* **23,** 1156-9.

Chan-Yeung, M., Ooi, G. C., Hui, D. S., Ho, P. L. & Tsang, K. W. (2003). Severe acute respiratory syndrome. *Int J Tuberc Lung Dis* **7,** 1117-30.

Chang, C. K., Sue, S. C., Yu, T. H., Hsieh, C. M., Tsai, C. K., Chiang, Y. C., Lee, S. J., Hsiao, H. H., Wu, W. J., Chang, W. L., Lin, C. H. & Huang, T. H. (2006). Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci* **13,** 59-72.

Chang, M. S., Lu, Y. T., Ho, S. T., Wu, C. C., Wei, T. Y., Chen, C. J., Hsu, Y. T., Chu, P. C., Chen, C. H., Chu, J. M., Jan, Y. L., Hung, C. C., Fan, C. C. & Yang, Y. C. (2004). Antibody detection of SARS-CoV spike and nucleocapsid protein. *Biochem Biophys Res Commun* **314,** 931-6.

Chau, T. N., Lee, K. C., Yao, H., Tsang, T. Y., Chow, T. C., Yeung, Y. C., Choi, K. W., Tso, Y. K., Lau, T., Lai, S. T. & Lai, C. L. (2004). SARS-associated viral hepatitis caused by a novel coronavirus: report of three cases. *Hepatology* **39,** 302-10.

Chaudhuri, T. K. & Gupta, P. (2005). Factors governing the substrate recognition by GroEL chaperone: a sequence correlation approach. *Cell Stress Chaperones* **10,** 24-36.

Chen, J., Miao, L., Li, J. M., Li, Y. Y., Zhu, Q. Y., Zhou, C. L., Fang, H. Q. & Chen, H. P. (2005). Receptor-binding domain of SARS-Cov spike protein: soluble expression in *E. coli*, purification and functional characterization. *World J Gastroenterol* **11,** 6159-64.

Chow, K. C., Hsiao, C. H., Lin, T. Y., Chen, C. L. & Chiou, S. H. (2004). Detection of severe acute respiratory syndrome-associated coronavirus in pneumocytes of the lung. *Am J Clin Pathol* **121,** 574-80.

Chow, K. Y., Hon, C. C., Hui, R. K., Wong, R. T., Yip, C. W., Zeng, F. & Leung, F. C. (2003). Molecular advances in severe acute respiratory syndrome-associated coronavirus (SARS-CoV). *Genomics Proteomics Bioinformatics* **1,** 247-62.

Chow, K. Y., Yeung, Y. S., Hon, C. C., Zeng, F., Law, K. M. & Leung, F. C. (2005). Adenovirus-mediated expression of the C-terminal domain of SARS-CoV spike protein is sufficient to induce apoptosis in Vero E6 cells. *FEBS Lett* **579,** 6699-704.

Choy, W. Y., Lin, S. G., Chan, P. K., Tam, J. S., Lo, Y. M., Chu, I. M., Tsai, S. N., Zhong, M. Q., Fung, K. P., Waye, M. M., Tsui, S. K., Ng, K. O., Shan, Z. X., Yang, M., Wu, Y. L., Lin, Z. Y. & Ngai, S. M. (2004). Synthetic peptide studies on the severe acute respiratory syndrome (SARS) coronavirus spike glycoprotein: perspective for SARS vaccine development *Clin Chem* **50,** 1036-42.

Cinatl, J. Jr., Michaelis, M., Hoever, G., Preiser, W. & Doerr, H. W. (2005). Development of antiviral therapy for severe acute respiratory syndrome. *Antiviral Res* **66,** 81-97.

Crackower, M. A., Sarao, R., Oudit, G. Y., Yagil, C., Kozieradzki, I., Scanga, S. E., Oliveira-dos-Santos, A. J., da Costa, J., Zhang, L., Pei, Y., Scholey, J., Ferrario, C. M., Manoukian, A. S., Chappell, M. C., Backx, P. H., Yagil, Y. & Penninger, J. M. (2002). Angiotensin-converting enzyme 2 is an essential regulator of heart function *Nature* **417,** 822-8.

Cregg, J. M., Vedvick, T. S. & Raschke, W. C. (1993). Recent advances in the expression of foreign genes in *Pichia pastoris*. *Biotechnology (N Y)* **11,** 905-10.

Daly, R. & Hearn, M. T. (2005). Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit* **18,** 119-38.

Di Guan, C., Li, P., Riggs, P. D. & Inouye, H. (1988). Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **67,** 21-30.

Ding, Y., He, L., Zhang, Q., Huang, Z., Che, X., Hou, J., Wang, H., Shen, H., Qiu, L., Li, Z., Geng, J., Cai, J., Han, H., Li, X., Kang, W., Weng, D., Liang, P. & Jiang, S. (2004). Organ distribution of severe acute respiratory syndrome (SARS) associated coronavirus (SARS-CoV) in SARS patients: implications for pathogenesis and virus transmission pathways. *J Pathol* **203,** 622-30.

Donnelly, C. A., Ghani, A. C., Leung, G. M., Hedley, A. J., Fraser, C., Riley, S., Abu-Raddad, L. J., Ho, L. M., Thach, T. Q., Chau, P., Chan, K. P., Lam, T. H., Tse, L. Y., Tsang, T., Liu, S. H., Kong, J. H., Lau, E. M., Ferguson, N. M., & Anderson, R. M. (2003). Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* **361,** 1761-6.

Donoghue, M., Hsieh, F., Baronas, E., Godbout, K., Gosselin, M., Stagliano, N., Donovan, M., Woolf, B., Robison, K., Jeyaseelan, R., Breitbart, R. E. & Acton, S. (2000). A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9.*Circ Res* **87,** E1-9.

Douglas, G. C., O'Bryan, M. K., Hedger, M. P., Lee, D. K., Yarski, M. A., Smith, A. I. & Lew, R. A. (2004). The novel angiotensin-converting enzyme (ACE) homolog, ACE2, is selectively expressed by adult Leydig cells of the testis. *Endocrinology* **145,** 4703-11.

Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A., Berger, A., Burguiere, A. M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J. C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H. D., Osterhaus, A. D., Schmitz & H., Doerr, H. W. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* **348,** 1967-76.

Dumon-Seignovert, L., Cariot, G. & Vuillard, L. (2004). The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif* **37,** 203-6.

Duquerroy, S., Vigouroux, A., Rottier, P. J., Rey, F. A. & Bosch, B. J. (2005). Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein. *Virology* **335,** 276-85.

Eckert, D. M. & Kim, P. S. (2001). Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem* **70,** 777-810.

Endo, Y., Nagai, H., Watanabe, Y., Ochi, K. & Takagi, T. (1992). Heat-induced aggregation of recombinant erythropoietin in the intact and deglycosylated states as monitored by gel permeation chromatography combined with a low-angle laser light scattering technique. *J Biochem (Tokyo)* **112,** 700-6.

Eriksson, U., Danilczyk, U. & Penninger, J. M. (2002). Just the beginning: novel functions for angiotensin-converting enzymes. *Curr Biol* **12,** 745-52.

Ewalt, K. L., Hendrick, J. P., Houry, W. A. & Hartl, F. U. (1997). In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell* **90,** 491-500.

Fang, X., Ye, L., Timani, K. A., Li, S., Zen, Y., Zhao, M., Zheng, H. & Wu, Z. (2005). Peptide domain involved in the interaction between membrane protein and nucleocapsid protein of SARS-associated coronavirus. *J Biochem Mol Biol* **38,** 381-5.

Feltham, J. L. & Gierasch, L. M. (2000). GroEL-substrate interactions: molding the fold, or folding the mold? *Cell* **100,** 193-6.

Flynn, G. C., Pohl, J., Flocco, M. T. & Rothman, J. E. (1991). Peptide-binding specificity of the molecular chaperone BiP. *Nature* **353,** 726-30.

Follis, K. E., York, J. & Nunberg, J. H. (2006). Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. *Virology* **350,** 358-69.

Fouchier, R. A., Kuiken, T., Schutten, M., van Amerongen, G., van Doornum, G. J., van den Hoogen, B. G., Peiris, M., Lim, W., Stohr, K. & Osterhaus, A. D. (2003). Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* **423,** 240.

Gao, F., Ou, H. Y., Chen, L. L., Zheng, W. X. & Zhang, C. T. (2003a). Prediction of proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes. *FEBS Lett* **553,** 451-456.

Gao, W., Tamin, A., Soloff, A., D'Aiuto, L., Nwanegbo, E., Robbins, P. D., Bellini, W. J., Barratt-Boyes, S. & Gambotto, A. (2003b). Effects of a SARS-associated coronavirus vaccine in monkeys *Lancet* **362,** 1895-6.

Gemmill, T. R. & Trimble, R. B. (1999). Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochim Biophys Acta* **1426,** 227-37.

Gibbs, A. J., Gibbs, M. J. & Armstrong, J. S. (2004). The phylogeny of SARS coronavirus. *Arch Virol* **149,** 621-4.

Giroglou, T., Cinatl, J. Jr., Rabenau, H., Drosten, C., Schwalbe, H., Doerr, H. W. & von Laer, D. (2004). Retroviral vectors pseudotyped with severe acute respiratory syndrome coronavirus S protein. *J Virol* **78,** :9007-15.

Gorbalenya, A. E., Snijder, E. J. & Spaan, W. J. (2004). Severe acute respiratory syndrome coronavirus phylogeny: toward consensus. *J Virol* **78,** 7863-6.

Grantcharova, V., Alm, E. J., Baker, D. & Horwich, A. L. (2001). Mechanisms of protein folding. *Curr Opin Struct Biol* **11,** 70-82.

Hakansson, K., Broder, D., Wang, A. H. & Miller, C. G. (2000). Crystallization of peptidase T from Salmonella typhimurium. *Acta Crystallogr D Biol Crystallogr* **56,** 924-6.

Hakansson-McReynolds, S., Jiang, S., Rong, L. & Caffrey, M. (2006). Solution structure of the SARS-coronavirus HR2 domain in the prefusion state. *J Biol Chem* **281,** 11965-71.

Hamilton, S. R., O'Donnell, J. B. Jr., Hammet, A., Stapleton, D., Habinowski, S. A., Means, A. R., Kemp, B. E. & Witters, L. A. AMP-activated protein kinase kinase: detection with recombinant AMPK alpha1 subunit. *Biochem Biophys Res Commun* **293,** 892-8.

Hamming, I., Timens, W., Bulthuis, M. L., Lely, A. T., Navis, G. J. & van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol* **203,** 631-7.

Harmer, D., Gilbert, M., Borman, R. & Clark, K. L. (2002). Quantitative mRNA expression profiling of ACE 2, a novel homologue of angiotensin converting enzyme. *FEBS Lett* **532,** 107-10.

Haas, I. G. (1994). BiP (GRP78), an essential hsp70 resident protein in the endoplasmic reticulum. *Experientia* **50,** 1012-20.

Haas, I. G. & Wabl, M. (1983). Immunoglobulin heavy chain binding protein. *Nature* **306,** 387-9.

He, R., Leeson, A., Ballantine, M., Andonov, A., Baker, L., Dobie, F., Li, Y., Bastien, N., Feldmann, H., Strocher, U., Theriault, S., Cutts, T., Cao, J., Booth, T. F., Plummer, F. A., Tyler, S. & Li, X. (2004). Characterization of protein-protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. *Virus Res* **105,** 121-5.

He, Y., Li, J. & Jiang, S. (2006). A single amino acid substitution (R441A) in the receptor-binding domain of SARS coronavirus spike protein disrupts the antigenic structure and binding activity. *Biochem Biophys Res Commun* **344,** 106-13.

Helenius, A. (1994). How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol Biol Cell* **5,** 253-65.

Ho, T. Y., Wu, S. L., Cheng, S. E., Wei, Y. C., Huang, S. P. & Hsiang, C. Y. (2004a). Antigenicity and receptor-binding ability of recombinant SARS coronavirus spike protein. *Biochem Biophys Res Commun* **313,** 938-47.

Ho, Y., Lin, P. H., Liu, C. Y., Lee, S. P. & Chao, Y. C. (2004b). Assembly of human severe acute respiratory syndrome coronavirus-like particles. *Biochem Biophys Res Commun* **318,** 833-8.

Hochuli, E., Dobeli, H. & Schacher, A. (1987). New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J Chromatogr* **411,** 177-84.

Hofmann, H., Geier, M., Marzi, A., Krumbiegel, M., Peipp, M., Fey, G. H., Gramberg, T. & Pohlmann, S. (2004a). Susceptibility to SARS coronavirus S protein-driven infection correlates with expression of angiotensin converting enzyme 2 and infection can be blocked by soluble receptor. *Biochem Biophys Res Commun* **319,** 1216-21.

Hofmann, H., Hattermann, K., Marzi, A., Gramberg, T., Geier, M., Krumbiegel, M., Kuate, S., Uberla, K., Niedrig, M. & Pohlmann, S. (2004b). S protein of severe acute respiratory syndrome-associated coronavirus mediates entry into hepatoma cell lines and is targeted by neutralizing antibodies in infected patients. *J Virol* **78,** 6134-42.

Hofmann, H. & Pohlmann, S. (2004c). Cellular entry of the SARS coronavirus. *Trends Microbiol* **12,** 466-72.

Horwich, A. L., Low, K. B., Fenton, W. A., Hirshfield, I. N. & Furtak, K. (1993). Folding in vivo of bacterial cytoplasmic proteins: role of GroEL. *Cell* **74,** 909-17.

Horzinek, M. C. (1999). Molecular evolution of corona- and toroviruses. *Adv Exp Med Biol* **473,** 61-72.

Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F. & Hartl, F. U. (1999). Identification of in vivo substrates of the chaperonin GroEL. *Nature* **402,** 147-54.

Hsieh, P. K., Chang, S. C., Huang, C. C., Lee, T. T., Hsiao, C. W., Kou, Y. H., Chen, I. Y., Chang, C. K., Huang, T. H. & Chang, M. F. (2005). Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. *J Virol* **79,** 13848-55.

Hsu, C. H., Ko, T. P., Yu, H. M., Tang, T. K., Chen, S. T. & Wang, A. H. (2004). Immunological, structural, and preliminary X-ray diffraction characterizations of the fusion core of the SARS-coronavirus spike protein. *Biochem Biophys Res Commun* **324,** 761-7.

Hu, S., Li, L., Qiao, J., Guo, Y., Cheng, L. & Liu, J. (2006). Codon optimization, expression, and characterization of an internalizing anti-ErbB2 single-chain antibody in *Pichia pastoris*. *Protein Expr Purif* **47,** 249-57.

Huang, Q., Yu, L., Petros, A. M., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S. W. & Olejniczak, E. T. (2004). Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry* **43,** 6059-63.

Ingallinella, P., Bianchi, E., Finotto, M., Cantoni, G., Eckert, D. M., Supekar, V. M., Bruckmann, C., Carfi, A. & Pessi, A. (2004). Structural characterization of the fusion-active complex of severe acute respiratory syndrome (SARS) coronavirus. *Proc Natl Acad Sci U S A* **101,** 8709-14.

Jafari-Aghdam, J., Khajeh, K., Ranjbar, B. & Nemat-Gorgani, M. (2005). Deglycosylation of glucoamylase from Aspergillus niger: effects on structure, activity and stability. *Biochim Biophys Acta* **1750,** 61-8.

Jeffers, S. A., Tusell, S. M., Gillim-Ross, L., Hemmila, E. M., Achenbach, J. E., Babcock, G. J., Thomas, W. D. Jr., Thackray, L. B., Young, M. D., Mason, R. J., Ambrosino, D. M., Wentworth, D. E., Demartini, J. C. & Holmes, K. V. (2004). CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A* **101,** 15748-53.

Kim, O. J., Lee, D. H. & Lee, C. H. (2006). Close relationship between SARS-coronavirus and group 2 coronavirus. *J Microbiol* **44,** 83-91.

Kotula, L. & Curtis, P. J. (1991). Evaluation of foreign gene codon optimization in yeast: expression of a mouse IG kappa chain. *Biotechnology (NY)* **9,** 1386-9.

Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., DeRisi, J., Yang J. Y., Cox, N., Hughes, J. M., LeDuc, J. W., Bellini, W. J., Anderson, L. J. & SARS Working Group. (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* **348,** 1953-66.

Kuiken, T., Fouchier, R. A., Schutten, M., Rimmelzwaan, G. F., van Amerongen, G., van Riel, D., Laman, J. D., de Jong, T., van Doornum, G., Lim, W., Ling, A. E., Chan, P. K., Tam, J. S., Zambon, M. C., Gopal, R., Drosten, C., van der Werf, S., Escriou, N., Manuguerra, J. C., Stohr, K., Peiris, J. S. & Osterhaus, A. D. (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* **362,** 263-70.

Kuo, L., Godeke, G. J., Raamsman, M. J., Masters, P. S. & Rottier, P. J. (2000). Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *J Virol* **74,** 1393-406.

Lee, J. S., Poo, H., Han, D. P., Hong, S. P., Kim, K., Cho, M. W., Kim, E., Sung, M. H. & Kim, C. J. (2006). Mucosal immunization with surface-displayed severe acute respiratory syndrome coronavirus spike protein on Lactobacillus casei induces neutralizing antibodies in mice *J Virol* **80,** 4079-87.

Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G. M., Ahuja, A., Yung, M. Y., Leung, C. B., To, K. F., Lui, S. F., Szeto, C. C., Chung, S. & Sung, J. J. (2003). A major outbreak of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* **348,** 1986-94.

Lee, Y. N., Chen, L. K., Ma, H. C., Yang, H. H., Li, H. P. & Lo, S. Y. (2005). Thermal aggregation of SARS-CoV membrane protein. *J Virol Methods* **129,** 152-61.

Lemonick, M. D. & Park, A. (2003, May 5). The truth about SARS *Time*, **161.**

Leung, W. K., To, K. F., Chan, P. K., Chan, H. L., Wu, A. K., Lee, N., Yuen, K. Y. & Sung, J. J. (2003). Enteric involvement of severe acute respiratory syndrome-associated coronavirus infection. *Gastroenterology* **125,** 1011-7.

Li, F., Li, W., Farzan, M. & Harrison, S. C. (2005a). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309,** 1864-8.

Li, T., Zhang, Y., Fu, L., Yu, C., Li, X., Li, Y., Zhang, X., Rong, Z., Wang, Y., Ning, H., Liang, R., Chen, W., Babiuk, L. A. & Chang, Z. (2005b). siRNA targeting the leader sequence of SARS-CoV inhibits virus replication. *Gene Ther* **12,** 751-61.

Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H. & Farzan, M. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426,** 450-4.

Liao, H. H. (1991). Effect of temperature on the expression of wild-type and thermostable mutants of kanamycin nucleotidyltransferase in *Escherichia coli*. *Protein Expr Purif* **2,** 43-50.

Liu, S., Xiao, G., Chen, Y., He, Y., Niu, J., Escalante, C. R., Xiong, H., Farmar, J., Debnath, A. K., Tien, P. & Jiang, S. (2004). Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet* **363,** 938-47.

Luo, H., Wu, D., Shen, C., Chen, K., Shen, X. & Jiang, H. (2006). Severe acute respiratory syndrome coronavirus membrane protein interacts with nucleocapsid protein mostly through their carboxyl termini by electrostatic attraction. *Int J Biochem Cell Biol* **38,** 589-99.

Manevich, Y., Feinstein, S. I. & Fisher, A. B. (2004). Activation of the antioxidant enzyme 1-CYS peroxiredoxin requires glutathionylation mediated by heterodimerization with pi GST. *Proc Natl Acad Sci U S A* **101,** 3780-5.

Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S., Khattra, J., Asano, J. K., Barber, S. A., Chan, S. Y., Cloutier, A., Coughlin, S. M., Freeman, D., Girn, N., Griffith, O. L., Leach, S. R., Mayo, M., McDonald, H., Montgomery, S. B., Pandoh, P. K., Petrescu, A. S., Robertson, A. G., Schein, J. E., Siddiqui, A., Smailus, D. E., Stott, J. M., Yang, G. S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T. F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G. A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R. C., Krajden, M., Petric, M., Skowronski, D. M., Upton, C. & Roper, R. L. (2003). The Genome sequence of the SARS-associated coronavirus. *Science* **300,** 1399-4.

Marzi, A., Gramberg, T., Simmons, G., Moller, P., Rennekamp, A. J., Krumbiegel, M., Geier, M., Eisemann, J., Turza, N., Saunier, B., Steinkasserer, A., Becker, S., Bates, P., Hofmann, H. & Pohlmann, S. (2004). DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. *J Virol* **78,** 12090-5.

Miroux, B. & Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260,** 289-98.

Molhoj, M., Ulvskov, P. & Dal Degan, F. (2001). Characterization of a functional soluble form of a Brassica napus membrane-anchored endo-1,4-beta-glucanase heterologously expressed in *Pichia pastoris*. *Plant Physiol* **127,** 674-84.

Moore, M. J., Dorfman, T., Li, W., Wong, S. K., Li, Y., Kuhn, J. H., Coderre, J., Vasilieva, N., Han, Z., Greenough, T. C., Farzan, M. & Choe, H. (2004). Retroviruses pseudotyped with the severe acute respiratory syndrome coronavirus spike protein efficiently infect cells expressing angiotensin-converting enzyme 2. *J Virol* **78,** 10628-35.

Ng, M. L., Tan, S. H., See, E. E., Ooi, E. E. & Ling, A. E. (2003a). Early events of SARS coronavirus infection in vero cells. *J Med Virol* 71, 323-31.

Ng, M. L., Tan, S. H., See, E. E., Ooi, E. E. & Ling, A. E. (2003b). Proliferative growth of SARS coronavirus in Vero E6 cells. *J Gen Virol* **84,** 3291-303.

Nie, Y., Wang, P., Shi, X., Wang, G., Chen, J., Zheng, A., Wang, W., Wang, Z., Qu, X., Luo, M., Tan, L., Song, X., Yin, X., Chen, J., Ding, M. & Deng, H. Highly infectious SARS-CoV pseudotyped virus reveals the cell tropism and its correlation with receptor expression. *Biochem Biophys Res Commun* **321,** 994-1000.

Ohnishi, K., Sakaguchi, M., Kaji, T., Akagawa, K., Taniyama, T., Kasai, M., Tsunetsugu-Yokota, Y., Oshima, M., Yamamoto, K., Takasuka, N., Hashimoto, S., Ato, M., Fujii, H., Takahashi, Y., Morikawa, S., Ishii, K., Sata, T., Takagi, H., Itamura, S., Odagiri, T., Miyamura, T., Kurane, I., Tashiro, M., Kurata, T., Yoshikura, H. & Takemori, T. (2005). Immunological detection of severe acute respiratory syndrome coronavirus by monoclonal antibodies. *Jpn J Infect Dis* **58,** 88-94.

Peiris, J. S., Chu, C. M., Cheng, V. C., Chan, K. S., Hung, I. F., Poon, L. L., Law, K. I., Tang, B. S., Hon, T. Y., Chan, C. S., Chan, K. H., Ng, J. S., Zheng, B. J., Ng, W. L., Lai, R. W., Guan, Y., Yuen, K. Y & HKU/UCH SARS Study Group. (2003a). Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361,** 1767-72.

Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W., Nicholls, J., Yee, W. K., Yan, W. W., Cheung, M. T., Cheng, V. C., Chan, K. H., Tsang, D. N., Yung, R. W., Ng, T. K., Yuen, K. Y. & SARS study group. (2003b). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361,** 1319-25.

Petit, C. M., Melancon, J. M., Chouljenko, V. N., Colgrove, R., Farzan, M., Knipe, D. M. & Kousoulas, K. G. (2005). Genetic analysis of the SARS-coronavirus spike glycoprotein functional domains involved in cell-surface expression and cell-to-cell fusion. *Virology* **341,** 215-30.

Phillips, J. J., Chua, M. M., Lavi, E. & Weiss, S. R. (1999). Pathogenesis of chimeric MHV4/MHV-A59 recombinant viruses: the murine coronavirus spike protein is a major determinant of neurovirulence. *J Virol* **73,** 7752-60.

Podmore, A. H. & Reynolds, P. E. (2002). Purification and characterization of VanXY(C), a D,D-dipeptidase/D,D-carboxypeptidase in vancomycin-resistant Enterococcus gallinarum BM4174. *Eur J Biochem* **269,** 2740-6.

Poutanen, S. M., Low, D. E., Henry, B., Finkelstein, S., Rose, D., Green, K., Tellier, R., Draker, R., Adachi, D., Ayers, M., Chan, A. K., Skowronski, D. M., Salit, I., Simor, A. E., Slutsky, A. S., Doyle, P. W., Krajden, M., Petric, M., Brunham, R. C., McGeer, A. J., National Microbiology Laboratory, Canada; (2003). Canadian Severe Acute Respiratory Syndrome Study Team. Identification of severe acute respiratory syndrome in Canada. *N Engl J Med* **348,** 1995-2005.

Qin, E., Shi, H., Tang, L., Wang, C., Chang, G., Ding, Z., Zhao, K., Wang, J., Chen, Z., Yu, M., Si, B., Liu, J., Wu, D., Cheng, X., Yang, B., Peng, W., Meng, Q., Liu, B., Han, W., Yin, X., Duan, H., Zhan, D., Tian, L., Li, S., Wu, J., Tan, G., Li, Y., Li, Y., Liu, Y., Liu, H., Lv, F., Zhang, Y., Kong, X., Fan, B., Jiang, T., Xu, S., Wang, X., Li, C., Wu, X., Deng, Y., Zhao, M. & Zhu, Q. (2006). Immunogenicity and protective efficacy in monkeys of purified inactivated Vero-cell SARS vaccine. *Vaccine* **24,** 1028-34.

Ralat, L. A., Manevich, Y., Fisher, A. B. & Colman, R. F. (2006). Direct evidence for the formation of a complex between 1-cysteine peroxiredoxin and glutathione S-transferase pi with activity changes in both enzymes. *Biochemistry* **45,** 360-72.

Reilley, B., Van Herp, M., Sermand, D. & Dentico, N. (2003). SARS and Carlo Urbani. *N Engl J Med* **348,** 1951-2.

Robertson, M. P., Igel, H., Baertsch, R., Haussler, D., Ares, M. Jr. and Scott, W. G. (2005). The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol* **3,** e5.

Romanos, M. A., Scorer, C. A. & Clare, J. J. (1992). Foreign gene expression in yeast: a review. *Yeast* **8,** 423-88.

Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Penaranda, S., Bankamp, B., Maher, K., Chen, M. H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J. L., Chen, Q., Wang, D., Erdman, D. D., Peret, T. C., Burns, C., Ksiazek, T. G., Rollin, P. E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A. D., Drosten, C., Pallansch, M. A., Anderson, L. J. & Bellini, W. J. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300,** 1394-9.

Sachdev, D. & Chirgwin, J. M. (1999). Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein. *J Protein Chem* **18,** 127-36.

Saibil, H. R. & Ranson, N. A. (2002). The chaperonin folding machine. *Trends Biochem Sci* **27,** 627-32.

Sainz, B. Jr., Mossel, E. C., Gallaher, W. R., Wimley, W. C., Peters, C. J., Wilson, R. B. & Garry, R. F. (2006). Inhibition of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) infectivity by peptides analogous to the viral spike protein. *Virus Res* **120,** 146-55.

Sainz, B. Jr., Rausch, J. M., Gallaher, W. R., Garry, R. F. & Wimley, W. C. (2005). Identification and characterization of the putative fusion peptide of the severe acute respiratory syndrome-associated coronavirus spike protein. *J Virol* **79,** 7195-206.

Sakikawa, C., Taguchi, H., Makino, Y. & Yoshida, M. (1999). On the maximum size of proteins to stay and fold in the cavity of GroEL underneath GroES. *J Biol Chem* **274,** 21251-6.

Sanchez, C. M., Izeta, A., Sanchez-Morgado, J. M., Alonso, S., Sola, I., Balasch, M., Plana-Duran, J. & Enjuanes, L. (1999). Targeted recombination demonstrates that the spike gene of transmissible gastroenteritis coronavirus is a determinant of its enteric tropism and virulence. *J Virol* **73,** 7607-18.

Sigler, P. B., Xu, Z., Rye, H. S., Burston, S. G., Fenton, W. A. & Horwich, A. L. (1998). Structure and function in GroEL-mediated protein folding. *Annu Rev Biochem* **67,** 581-608.

Simmons, G., Reeves, J. D., Rennekamp, A. J., Amberg, S. M., Piefer, A. J. & Bates, P. (2004). Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proc Natl Acad Sci U S A* **101,** 4240-5.

Smith, A. E. & Helenius, A. (2004). How viruses enter animal cells. *Science* **304,** 237-42.

Smith, D. B. & Johnson, K. S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67,** 31-40.

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J. & Gorbalenya, A. E. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol* **331,** 991-1004.

Song, H. C., Seo, M. Y., Stadler, K., Yoo, B. J., Choo, Q. L., Coates, S. R., Uematsu, Y., Harada, T., Greer, C. E., Polo, J. M., Pileri, P., Eickmann, M., Rappuoli, R., Abrignani, S., Houghton, M. & Han, J. H. (2004). Synthesis and characterization of a native, oligomeric form of recombinant severe acute respiratory syndrome coronavirus spike glycoprotein. *J Virol* **78,** 10328-35.

Sreekrishna, K., Brankamp, R. G., Kropp, K. E., Blankenship, D. T., Tsay, J. T., Smith, P. L., Wierschke, J. D., Subramaniam, A. & Birkenberger, L. A. (1997). Strategies for optimal synthesis and secretion of heterologous proteins in the methylotrophic yeast *Pichia pastoris*. *Gene* **190,** 55-62.

Stan, G., Brooks, B. R., Lorimer, G. H. & Thirumalai, D. (2006). Residues in substrate proteins that interact with GroEL in the capture process are buried in the native state. *Proc Natl Acad Sci U S A* **103,** 4433-8.

Strandberg, L. & Enfors, S. O. (1991). Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*. *Appl Environ Microbiol* **57,** 1669-74.

Sui, J., Li, W., Murakami, A., Tamin, A., Matthews, L. J., Wong, S. K., Moore, M. J., Tallarico, A. S., Olurinde, M., Choe, H., Anderson, L. J., Bellini, W. J., Farzan, M. & Marasco, W. A. (2004). Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc Natl Acad Sci U S A* **101,** 2536-41.

Supekar, V. M., Bruckmann, C., Ingallinella, P., Bianchi, E., Pessi, A. & Carfi, A. (2004). Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus S2 fusion protein. *Proc Natl Acad Sci U S A* **101,** 17958-63.

Taguchi, H. (2005). Chaperonin GroEL meets the substrate protein as a "load" of the rings. *J Biochem (Tokyo)* **137,** 543-9.

Tarentino, A. L., Gomez, C. M. & Plummer, T. H. Jr. (1985). Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry* **24,** 4665-71.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **60,** 523-3.

Thiel, V., Ivanov, K. A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E. & Ziebuhr, J. (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol* **84,** 2305-15.

Thirumalai, D. & Lorimer, G. H. (2001). Chaperonin-mediated protein folding. *Annu Rev Biophys Biomol Struct* **30,** 245-69.

Tikellis, C., Johnston, C. I., Forbes, J. M., Burns, W. C., Burrell, L. M., Risvanis, J. & Cooper, M. E. (2003). Characterization of renal angiotensin-converting enzyme 2 in diabetic nephropathy. *Hypertension* **41,** 392-7.

Tipnis, S. R., Hooper, N. M., Hyde, R., Karran, E., Christie, G. & Turner, A. J. (2000). A human homolog of angiotensin-converting enzyme. Cloning and functional expression as a captopril-insensitive carboxypeptidase. *J Biol Chem* **275,** 33238-43.

To, K. F. & Lo, A. W. (2004a). Exploring the pathogenesis of severe acute respiratory syndrome (SARS): the tissue distribution of the coronavirus (SARS-CoV) and its putative receptor, angiotensin-converting enzyme 2 (ACE2). *J Pathol* **203,** 740-3.

To, K. F., Tong, J. H., Chan, P. K., Au, F. W., Chim, S. S., Chan, K. C., Cheung, J. L., Liu, E. Y., Tse, G. M., Lo, A. W., Lo, Y. M. & Ng, H. K. (2004b). Tissue and cellular tropism of the coronavirus associated with severe acute respiratory syndrome: an in-situ hybridization study of fatal cases. *J Pathol* **202,** 157-63..

Tripet, B., Howard, M. W., Jobling, M., Holmes, R. K., Holmes, K. V. & Hodges, R. S. (2004). Structural characterization of the SARS-coronavirus spike S fusion protein core. *J Biol Chem* **279,** 20836-49.

Tsang, K. W., Ho, P. L., Ooi, G. C., Yee, W. K., Wang, T., Chan-Yeung, M., Lam, W. K., Seto, W. H., Yam, L. Y., Cheung, T. M., Wong, P. C., Lam, B., Ip, M. S., Chan, J., Yuen, K. Y. & Lai, K. N., (2003). A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* **348,** 1977-1985.

Tse, G. M., To, K. F., Chan, P. K., Lo, A. W., Ng, K. C., Wu, A., Lee, N., Wong, H. C., Mak, S. M., Chan, K. F., Hui, D. S., Sung, J. J. & Ng, H. K. (2004). Pulmonary pathological features in coronavirus associated severe acute respiratory syndrome. *J Clin Pathol* **57,** 260-5.

Tsui, S. K., Chim, S. S. & Lo, Y. M. (2003). Chinese University of Hong Kong Molecular SARS Research Group. Coronavirus genomic-sequence variations and the epidemiology of the severe acute respiratory syndrome. *N Engl J Med* **349,** 187-8.

Van Dyk, T. K., Gatenby, A. A. & LaRossa, R. A. (1989). Demonstration by genetic suppression of interaction of GroE products with many proteins. *Nature* **342,** 451-3.

Venkatesh, B., Arifuzzaman, M., Mori, H., Taguchi, T. & Ohmiya, Y. (2004). GroEL chaperone binding to beetle luciferases and the implications for refolding when co-expressed. *Biosci Biotechnol Biochem* **68,** 2096-103.

Viitanen, P. V., Gatenby, A. A. & Lorimer, G. H. (1992). Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of *Escherichia coli* proteins. *Protein Sci* **1,** 363-9.

Wang, C., Eufemi, M., Turano, C. & Giartosio, A. (1996). Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry* **35,** 7299-307.

Wang, Z., Yuan, Z., Matsumoto, M., Hengge, U. R. & Chang, Y. F. (2005). Immune responses with DNA vaccines encoded different gene fragments of severe acute respiratory syndrome coronavirus in BALB/c mice. *Biochem Biophys Res Commun* **327,** 130-5.

Wei, H. Y., Wang, J. W., Ou, Y. J., Wang, Y. B., Qu, J. G., Zhao, W. M. & Hong, T. (2005). Receptor-binding ability of fragments 260-600 and 397-796 of SARS-associated coronavirus spike protein. *Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi* **19,** 353-7.

Wei, J. & Hendershot, L. M. (1996). Protein folding and assembly in the endoplasmic reticulum. *EXS* **77,** 41-55.

World Health Organization. (2003a). Global surveillance for severe acute respiratory syndrome (SARS). *Wkly Epidemiol Rec* **78,** 100-19.

World Health Organization. (2003b). Severe acute respiratory syndrome (SARS): Status of the outbreak and lessons for the immediate future. Retrieved May 20, 2003, from http://www.who.int/csr/media/sars_wha.pdf

World Health Organization. (2003c). Summary table of SARS cases by country, 1 November 2002 - 7 August 2003. Retrieved from http://www.who.int/csr/sars/en/WHOconsensus.pdf

Wong, S. K., Li, W., Moore, M. J., Choe, H. & Farzan, M. (2004). A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem* **279,** 3197-201.

Wu, J. & Filutowicz, M. (1999). Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. *Acta Biochim Pol* **46,** 591-9.

Wu, Q., Zhang, Y., Lu, H., Wang, J., He, X., Liu, Y., Ye, C., Lin, W., Hu, J., Ji, J., Xu, J., Ye, J., Hu, Y., Chen, W., Li, S., Wang, J., Wang, J., Bi, S. & Yang H. (2003). The E protein is a multifunctional membrane protein of SARS-CoV. *Genomics Proteomics Bioinformatics* **1,** 131-44.

Wu, W., Wang, J., Liu, P., Chen, W., Yin, S., Jiang, S., Yan, L., Zhan, J., Chen, X., Li, J., Huang, Z. & Huang, H. (2003). A hospital outbreak of severe acute respiratory syndrome in Guangzhou, China. *Chin Med J* **116,** 811-818.

Xiao, X., Chakraborti, S., Dimitrov, A. S., Gramatikoff, K. & Dimitrov, D. S. (2003). The SARS-CoV S glycoprotein: expression and functional characterization. *Biochem Biophys Res Commun* **312,** 1159-64.

Xiao, X., Feng, Y., Chakraborti, S. & Dimitrov, D. S. (2004). Oligomerization of the SARS-CoV S glycoprotein: dimerization of the N-terminus and trimerization of the ectodomain. *Biochem Biophys Res Commun* **322,** 93-9.

Xiong, A. S., Yao, Q. H., Peng, R. H., Han, P. L., Cheng, Z. M. & Li Y. (2005). High level expression of a recombinant acid phytase gene in *Pichia pastoris*. *J Appl Microbiol* **98,** 418-28.

Xiong, A. S., Yao, Q. H., Peng, R. H., Zhang, Z., Xu, F., Liu, J. G., Han, P. L. & Chen, J. M. (2006). High level expression of a synthetic gene encoding Peniophora lycii phytase in methylotrophic yeast *Pichia pastoris*. *Appl Microbiol Biotechnol*

Xu, Y., Lou, Z., Liu, Y., Pang, H., Tien, P., Gao, G. F. & Rao, Z. (2004a). Crystal structure of severe acute respiratory syndrome coronavirus spike protein fusion core. *J Biol Chem* **279,** 49414-9.

Xu, Y., Su, N., Qin, L., Bai, Z., Gao, G. F. & Rao, Z. (2004b). Crystallization and preliminary crystallographic analysis of the heptad-repeat complex of SARS coronavirus spike protein. *Acta Crystallogr D Biol Crystallogr* **60,** 2377-9.

Xu, Y., Zhu, J., Liu, Y., Lou, Z., Yuan, F., Liu, Y., Cole, D. K., Ni, L., Su, N., Qin, L., Li, X., Bai, Z., Bell, J. I., Pang, H., Tien, P., Gao, G. F. & Rao, Z. (2004c). Characterization of the heptad repeat regions, HR1 and HR2, and design of a fusion core structure model of the spike protein from severe acute respiratory syndrome (SARS) coronavirus. *Biochemistry* **43,** 14064-71.

Yang, Z. Y., Huang, Y., Ganesh, L., Leung, K., Kong, W. P., Schwartz, O., Subbarao, K. & Nabel, G. J. (2004a). pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN. *J Virol* **78,** 5642-50.

Yang, Z. Y., Kong, W. P., Huang, Y., Roberts, A., Murphy, B. R., Subbarao, K. & Nabel, G. J. (2004b). A DNA vaccine induces SARS coronavirus neutralization and protective immunity in mice. *Nature* **428,** 561-4.

Yeung, K. S., Yamanaka, G. A. & Meanwell, N. A. (2006). Severe acute respiratory syndrome coronavirus entry into host cells: Opportunities for therapeutic intervention. *Med Res Rev* **26,** 414-33.

Yu, H., Yang, Y., Zhang, W., Xie, Y. H., Qin, J., Wang, Y., Zheng, H. B., Zhao, G. P., Yang, S. & Jiang, W. H. (2003). Expression and purification of recombinant SARS coronavirus spike protein *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* **35,** 774-8.

Zeng, F., Chow, K. Y., Hon, C. C., Law, K. M., Yip, C. W., Chan, K. H., Peiris, J. S. & Leung, F. C. (2004). Characterization of humoral responses in mice immunized with plasmid DNAs encoding SARS-CoV spike gene fragments. *Biochem Biophys Res Commun* **315,** 1134-9.

Zeng, F. Y., Chan, C. W., Chan, M. N., Chen, J. D., Chow, K. Y., Hon, C. C., Hui, K. H., Li, J., Li, V. Y., Wang, C. Y., Wang, P. Y., Guan, Y., Zheng, B., Poon, L. L., Chan, K. H., Yuen, K. Y., Peiris, J. S. & Leung, F. C. (2003). The complete genome sequence of severe acute respiratory syndrome coronavirus strain HKU-39849 (HK-39). *Exp Biol Med* **228,** 866-73.

Zhang, Q. F., Cui, J. M., Huang, X. J., Lin, W., Tan, D. Y., Xu, J. W., Yang, Y. F., Zhang, J. Q., Zhang, X., Li, H., Zheng, H. Y., Chen, Q. X., Yan, X. G., Zheng, K., Wan, Z. Y. & Huang, J. C. (2003). Morphology and morphogenesis of severe acute respiratory syndrome (SARS)-associated virus. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao* **35,** 587-91.

Zhang, Q. F., Cui, J. M., Huang, X. J., Zheng, H. Y., Hung, J. C., Fang, L., Li, K. P. & Zhang, J. Q. (2004). The life cycle of SARS coronavirus in Vero E6 cells. *J Med Virol* **73,** 332-7.

Zhang, Y., Li, T., Fu, L., Yu, C., Li, Y., Xu, X., Wang, Y., Ning, H., Zhang, S., Chen, W., Babiuk, L. A. & Chang, Z. (2004). Silencing SARS-CoV Spike protein expression in cultured cells by RNA interference. *FEBS Lett* **560,** 141-6.

Zhao, C. H., Guo, Y. B., Wu, H., Li, X. H., Guo, X. H., Jin, R. H, Dign, H. G., Meng, Q. H., Lang, Z. W., Wang, W., Yan, H. P., Huang, C. & Liu, D. G. (2003a) Clinical manifestation, treatment, and outcome of severe acute respiratory syndrome: analysis of 108 cases in Beijing. *Zhonghua Yi Xue Za Zhi* **83,** 897-901.

Zhao, J. C., Zhao, Z. D., Wang, W. & Gao, X. M. (2005). Prokaryotic expression, refolding, and purification of fragment 450-650 of the spike protein of SARS-coronavirus. *Protein Expr Purif* **39,** 169-74.

Zhao, P., Ke, J. S., Qin, Z. L., Ren, H., Zhao, L. J., Yu, J. G., Gao, J., Zhu, S. Y. & Qi, Z. T. (2004). DNA vaccine of SARS-Cov S gene induces antibody response in mice. *Acta Biochim Biophys Sin (Shanghai)* **36,** 37-41.

Zhao, Z., Zhang, F., Xu, M., Huang, K., Zhong, W., Cai, W., Yin, Z., Huang, S., Deng, Z., Wei, M., Xiong, J.& Hawkey, P.M. (2003b). Description and clinical treatment of an early outbreak of severe acute respiratory syndrome in Guangzhou, PR China. *J Med Microbiol* **52,** 715-20.

Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, Xie, Z. H., Chan, K. H., Li, P. H., Tan, S. Y., Chang, Q., Xie, J. P., Liu, X. Q., Xu, J., Li, D. X., Yuen, K. Y., Peiris & Guan, Y. (2003). Epidemiology and cause of severe acute respiratory syndrome in Guangdong, People's Republic of China, in February, 2003. *Lancet* **362,** 1353-8.

Zhou, T., Wang, H., Luo, D., Rowe, T., Wang, Z., Hogan, R. J., Qiu, S., Bunzel, R. J., Huang, G., Mishra, V., Voss, T. G., Kimberly, R. & Luo, M. (2004). An exposed domain in the severe acute respiratory syndrome coronavirus spike protein induces neutralizing antibodies. *J Virol* **78,** 7217-26.

Zhou, Z., Post, P., Chubet, R., Holtz, K., McPherson, C., Petric, M. & Cox, M. (2006). A recombinant baculovirus-expressed S glycoprotein vaccine elicits high titers of SARS-associated coronavirus (SARS-CoV) neutralizing antibodies in mice. *Vaccine* **24,** 3624-31.

Zhu, J., Xiao, G., Xu, Y., Yuan, F., Zheng, C., Liu, Y., Yan, H., Cole, D. K., Bell, J. I., Rao, Z., Tien, P. & Gao, G. F. (2004). Following the rule: formation of the 6-helix bundle of the fusion core from severe acute respiratory syndrome coronavirus spike protein and identification of potent peptide inhibitors. *Biochem Biophys Res Commun* **319,** 283-8.

Zisman, L. S., Keller, R. S., Weaver, B., Lin, Q., Speth, R., Bristow, M. R. & Canver, C. C. Increased angiotensin-(1-7)-forming activity in failing human heart ventricles: evidence for upregulation of the angiotensin-converting enzyme Homologue ACE2. *Circulation* **108,** 1707-12

Zuo, X., Mattern, M. R., Tan, R., Li, S., Hall, J., Sterner, D. E., Shoo, J., Tran, H., Lim, P., Sarafianos, S. G., Kazi, L., Navas-Martin, S., Weiss, S. R. & Butt, T. R. (2005). Expression and purification of SARS coronavirus proteins using SUMO-fusions. *Protein Expr Purif* **42,** 100-10.